

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 13, 2004, 14:01:27 ; Search time 68.5 seconds

(Without alignments)
9697.360 Million cell updates/sec

Title: NP000123-328

Perfect score: 12416

Sequence: 1 MQELSTCFCLLRPFCSA.....WVHQIALMEVLGCEAQQDLY 2351

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A Geneseq 29Jan04:*

1: geneseqp1980s:*

2: geneseqp1990s:*

3: geneseqp2000s:*

4: geneseqp2001s:*

5: geneseqp2002s:*

6: geneseqp2003as:*

7: geneseqp2003bs:*

8: geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	12410	100.0	2351	1 AAP50059	Human fac
2	12410	100.0	2351	2 AAW11329	Native Fa
3	12410	100.0	2351	2 AAW46245	Human fac
4	12410	100.0	2351	2 AAW44373	Human fac
5	12410	100.0	2351	3 AAY52537	Human fac
6	12410	100.0	2351	6 ABB99528	Amino aci
7	12406	99.9	2351	2 AAW11416	Active Fa
8	12406	99.9	2351	2 AAW11343	Active Fa
9	12406	99.9	2351	2 AAW11435	Active Fa
10	12405	99.9	2351	2 AAW11445	Active Fa
11	12405	99.9	2351	2 AAW11398	Active Fa
12	12405	99.9	2351	2 AAW11425	Active Fa
13	12405	99.9	2351	2 AAW11362	Active Fa
14	12405	99.9	2351	2 AAW11461	Active Fa
15	12405	99.9	2351	2 AAW11387	Active Fa
16	12405	99.9	2351	2 AAW11419	Active Fa
17	12404	99.9	2351	1 AAP81113	Factor VI
18	12404	99.9	2351	1 AAP80659	Sequence
19	12404	99.9	2351	2 AAW11408	Active Fa
20	12404	99.9	2351	2 AAW11427	Active Fa
21	12404	99.9	2351	2 AAW11347	Active Fa
22	12404	99.9	2351	2 AAW11332	Active Fa
23	12404	99.9	2351	2 AAW11437	Active Fa
24	12403	99.9	2351	2 AAW11371	Active Fa
25	12403	99.9	2351	2 AAW11396	Active Fa

ALIGNMENTS

RESULT 1
AAP50059
ID AAP50059 standard; protein; 2351 AA.

XX AAP50059;

DT 27-OCT-1991 (first entry)

DE Human factor VIII.

XX Human factor VIII.

XX Homo sapiens.

XX EP160457-A.

PD 06-NOV-1985.

XX 18-APR-1985; 85EP-00302734.

XX 20-APR-1984; 84US-00602312.

XX (GETH) GENENTECH INC.

PA Capon DJ, Vehar GA, Lawn RM, Wood WI;

XX WFI: 1985-277976/45.

DR N-PSDB; AAN50054.

PT New recombinant functional human factor VIII or deriv. - useful for treating haemophilia and obtd. as pure prod. by recombinant DNA technology.

XX Disclosure; Fig 10a(10-10C(III)); 109pp; English.

PS The sequence is that of human factor VIII. Amino acids 1-19 are the predicted signal peptide, and amino acids 1-2332 are the predicted mature protein. The protein is produced in pure form and in useful amts. using recombinant DNA techniques. Factor VIII can be used to correct factor VIII deficient plasma, and activates factor X to Xa in the presence of factor IXa, Ca and phospholipid. These activities are inactivated by antibodies specific for factor VIII. The activity of the prod. is bound to an immobilised monoclonal antibody specific for factor VIII. Factor VIII activity is activated by thrombin. The activity is bound to, and can be eluted from, immobilised von Willebrand factor. Dose of factor VIII is 20-40 units/kg over 8 hours i.v. for maintenance therapy for haemophilia, 40 units/kg for preoperative conditions, or 15-20 units/kg for minor haemorrhaging

Hit wild type
Factor VIII

XX	SQ	Sequence	2351	AA;
		Query Match	100.0%;	Score 12410; DB 1; Length 2351;
		Best Local Similarity	100.0%;	Pred. No. 0;
		Matches 2350; Conservative	0;	Mismatches 1; Indels 0; Gaps 0;
Qy	1	MQIELSTCFPLCLLRFCFSATRIYVLGAVELSDWYMQSDLGELPVDARPPRVKPSFFPN	60	
Dd	1	MQIELSTCFPLCLLRFCFSATRIYVLGAVELSDWYMQSDLGELPVDARPPRVKPSFFPN	60	
Qy	61	TSVYVKTLFVEFTDHLFNIAKPRPPWMLGPTIQAEVYDTWVITLKNMASEPVSLSHAV	120	
Dd	61	TSVYVKTLFVEFTDHLFNIAKPRPPWMLGPTIQAEVYDTWVITLKNMASHPSLSHAV	120	
Qy	121	GVSYWKASCAEAYDDQTSQREKDDKVFPGSGSHYUQVLEKNGPMASDPLCLTYSYLSH	180	
Dd	121	GVSYWKASCAEAYDDQTSQREKDDKVFPGSGSHYUQVLEKNGPMASDPLCLTYSYLSH	180	
Qy	181	VDLVKDLNSGLIGALLVCRGSLAKETQTLHKFILLFAVDEGKSWHSETKNSLMQDRD	240	
Dd	181	VDLVKDLNSGLIGALLVCRGSLAKETQTLHKFILLFAVDEGKSWHSETKNSLMQDRD	240	
Qy	241	AASARAWPKVHTVNGYVNRSLPGLIGCHRSVYVHWI GMCTTPEVHSIFLEGHTFLVRNE	300	
Dd	241	AASARAWPKVHTVNGYVNRSLPGLIGCHRSVYVHWI GMCTTPEVHSIFLEGHTFLVRNH	300	
Qy	301	ROASLESPITFLTAQTLMLDGLQFLLSCHISSHOHDGMEAYVKVDCSCEPQLRMKNE	360	
Dd	301	ROASLESPITFLTAQTLMLDGLQFLLSCHISSHOHDGMEAYVKVDCSCEPQLRMKNE	360	
Qy	361	EAEDYDDDLTDSMDVVRFDJDNKSPSFIQRSVAKKPKTWIHYIAAEEDWDYAPLVLA	420	
Dd	361	EAEDYDDDLTDSMDVVRFDJDNKSPSFIQRSVAKKPKTWIHYIAAEEDWDYAPLVLA	420	
Qy	421	PDDRSYKSQYLNNGPQRIGRYKKVRFMAYDTDFTKTREAIQHESGILGPLLYGEVGDTL	480	
Dd	421	PDDRSYKSQYLNNGPQRIGRYKKVRFMAYDTDFTKTREAIQHESGILGPLLYGEVGDCL	480	
Qy	481	LIIFKNQASRPYNIYPHGITDVRPLYSRRLPKGVKHLKDFEILPGRIFKYKWTVTVEDE	540	
Dd	481	LIIFKNQASRPYNIYPHGITDVRPLYSRRLPKGVKHLKDFEILPGRIFKYKWTVTVEDE	540	
Qy	541	TKSDPRCLTRYSSGFVNRDLASGLTGPLLICVKEVDQGNQIMSKDNVILFSVFDE	600	
Dd	541	TKSDPRCLTRYSSGFVNRDLASGLTGPLLICVKEVDQGNQIMSKDNVILFSVFDE	600	
Qy	601	NRSWYLTENIQRELNPAGVOLEDPEFQASNIMHSINGYVDFSLQLSVCLHEVAYWYLLS	660	
Dd	601	NRSWYLTENIQRELNPAGVOLEDPEFQASNIMHSINGYVDFSLQLSVCLHEVAYWYLLS	660	
Qy	661	IGAQTDFLSVPSFGYTFKHKNVYEDTLTPPSPGETVPMSENPGLWILGCHNSDFNRNG	720	
Dd	661	IGAQTDFLSVPSFGYTFKHKNVYEDTLTPPSPGETVPMSENPGLWILGCHNSDFNRNG	720	
Qy	721	MTALLKVSSCDKNCTGDYVEDSVEDISAVLLSKNNNAIPRSPQNSRHPSTROKQFNATTI	780	
Dd	721	MTALLKVSSCDKNCTGDYVEDSVEDISAVLLSKNNNAIPRSPQNSRHPSTROKQFNATTI	780	
Qy	781	PENDIEKTDPMFAHRTPMFKIQNVSSSDLLMLLRQSPTPHGLSLSDIQEAKYTFSDDPDS	840	
Dd	781	PENDIEKTDPMFAHRTPMFKIQNVSSSDLLMLLRQSPTPHGLSLSDIQEAKYTFSDDPDS	840	
Qy	841	PGAISNNLSLSEWTHFRPQLHHSGDWVTPPESGLOLRNLKGLTAATELTKLDPKVSST	900	
Dd	841	PGAISNNLSLSEWTHFRPQLHHSGDWVTPPESGLOLRNLKGLTAATELTKLDPKVSST	900	
Qy	901	SNNLISTPDSNLAAGTNTSILGPSPVHYVDOLDTTLFGKXSSPLTSEGGLPSLSEE	960	
Dd	901	SNNLISTPDSNLAAGTNTSILGPSPVHYVDOLDTTLFGKXSSPLTSEGGLPSLSEE	960	
Qy	961	NNDSKLLESLGMNSQESGWKGVSSSTEGRIIPKGRHGPALLTKDNALFKVSI SLLKTN	1020	

961	Db	NNDSKLLESLMNSQESSWGKVVSTESGRIPFKGRAHGPALLTKDNALFKVYSILSLTKTN	1020
1021	Qy	KTSNNSATNRKTHIDGSPILLIENSFSVWQNILESDETEFKKVTPLIHDRMLDKNATAPRL	1080
1021	Db	KTSNNSATNRKTHIDGSPILLIENSFSVWQNILESDETEFKKVTPLIHDRMLDKNATAPRL	1080
1081	Qy	NHNSNKITTSKNMEXVQOKKEGPIPPCAQNPOMSPKMLFLPESARWIORTGKGSLSNG	1140
1081	Db	NHNSNKITTSKNMEXVQOKKEGPIPPCAQNPOMSPKMLFLPESARWIORTGKGSLSNG	1140
1141	Qy	QGPSPKOLVSLGPEKSVGEQNTSEKNKVVGKGEFTKDVGLKEMVFPSSRNLFJTNLDN	1200
1141	Db	QGPSPKOLVSLGPEKSVGEQNTSEKNKVVGKGEFTKDVGLKEMVFPSSRNLFJTNLDN	1200
1201	Qy	LHENNTHNQEKKIQEIEKXETLIQENVVLPOHTVTGTKNPMKNLFLIUSTKONVEGSDV	1260
1201	Db	LHENNTHNQEKKIQEIEKXETLIQENVVLPOHTVTGTKNPMKNLFLIUSTKONVEGSDV	1260
1261	Qy	GAYAPVLODPRSLNDSTNRTKGHTAHSFKKGBEENLEGLGNQTKQIVEKYACTTRISPNIT	1320
1261	Db	GAYAPVLODPRSLNDSTNRTKGHTAHSFKKGBEENLEGLGNQTKQIVEKYACTTRISPNIT	1320
1321	Qy	SQONFVTOBSKRALKQFRLPLESTELEKRIIVDDTSTQWSKNMKHLTPSTLTQIDYNEKE	1380
1321	Db	SQONFVTOBSKRALKQFRLPLESTELEKRIIVDDTSTQWSKNMKHLTPSTLTQIDYNEKE	1380
1381	Qy	KGAITQSPISDCLTRGHSIPQANRSPPLIAKVSFSPSIRPIYITRVLFQDNSSHLPAAASY	1440
1381	Db	KGAITQSPISDCLTRGHSIPQANRSPPLIAKVSFSPSIRPIYITRVLFQDNSSHLPAAASY	1440
1441	Qy	RKDSGVQBSHSHFLOGAKKNLSIALITLMTGDQREVSGSLGTSANSTVYKKVENVPLP	1500
1441	Db	RKDSGVQBSHSHFLOGAKKNLSIALITLMTGDQREVSGSLGTSANSTVYKKVENVPLP	1500
1501	Qy	KPDLPKTSKGVEJLPPKVHIYQKDLFPETTSNGSPGHLDLVEGSLLOQTEGAIKWNEANRP	1560
1501	Db	KPDLPKTSKGVEJLPPKVHIYQKDLFPETTSNGSPGHLDLVEGSLLOQTEGAIKWNEANRP	1560
1561	Qy	GKVPFLVAVATESSAKTESKLLDPLANDNHVGTQIPKEEMKSQKSPKTAFAKKKDTILSL	1620
1561	Db	GKVPFLVAVATESSAKTESKLLDPLANDNHVGTQIPKEEMKSQKSPKTAFAKKKDTILSL	1620
1621	Qy	NACSNHAIIAINQONKPELEVTTWAKQGRTERLCSQNPVLRKHQSEIIRTTILQSDOE	1680
1621	Db	NACSNHAIIAINQONKPELEVTTWAKQGRTERLCSQNPVLRKHQSEIIRTTILQSDOE	1680
1681	Qy	IDYDDTISVENMKKEDFDIYEDENQSPRSFOKTRHYFIAAVERLWDYGSSSPHVLNRR	1740
1681	Db	IDYDDTISVENMKKEDFDIYEDENQSPRSFOKTRHYFIAAVERLWDYGSSSPHVLNRR	1740
1741	Qy	AQSGSVPOFKKVVQEFITDGSFTQPLYRGELNEHLGLGPYIRAEVSDNIMVTFRNOQASR	1800
1741	Db	AQSGSVPOFKKVVQEFITDGSFTQPLYRGELNEHLGLGPYIRAEVSDNIMVTFRNOQASR	1800
1801	Qy	PYSFYSSLSIYEEBQRCABPRKQFVPKNETKTYFWKVQHMAETKQEFDCKAWAYPSDV	1860
1801	Db	PYSFYSSLSIYEEBQRCABPRKQFVPKNETKTYFWKVQHMAETKQEFDCKAWAYPSDV	1860
1861	Qy	DLEKDVHSLGLPLLVCHTNTLNAHGRQVTVQEFALFTTIFDTSKGWYFTENVERNCR	1920
1861	Db	DLEKDVHSLGLPLLVCHTNTLNAHGRQVTVQEFALFTTIFDTSKGWYFTENVERNCR	1920
1921	Qy	PCNIQWEDPTPKENYRFAHNGYIMDTLPLGVMAQDQIRWYLLSMGSENENIHSIFSGH	1980
1921	Db	PCNIQWEDPTPKENYRFAHNGYIMDTLPLGVMAQDQIRWYLLSMGSENENIHSIFSGH	1980
1981	Qy	VFTVRKXEYXOALVNLVPGVFVTEVMIPSKAGIWRVBECLIGEHLHAGMSTLFLVYSNKC	2040
1981	Db	VFTVRKXEYXOALVNLVPGVFVTEVMIPSKAGIWRVBECLIGEHLHAGMSTLFLVYSNKC	2040
2041	Qy	QTPPLGMASGHIDFOITASGOYQWAPKLABLPHVSGSINASTKEPSTWIKVDLAPMII	2100
2041	Db	QTPPLGMASGHIDFOITASGOYQWAPKLABLPHVSGSINASTKEPSTWIKVDLAPMII	2100

CC that ultimately converts soluble fibrinogen to insoluble fibrin clot, effecting haemostasis. A deficiency in factor VIII:C is responsible for haemophilia A, which is an X-chromosome-linked inherited bleeding diathesis. Factor VIII:C is activated by plasma proteases, such as thrombin. During activation the mature polypeptide is cleaved to generate heavy and light chain fragments that are further cleaved. The DNA encoding this sequence is mutated, using mutagenic primers, to produce the active Factor VIII:C analogues of the invention (such as AAW11330). The analogues comprise a native Factor VIII:C polypeptide modified at a site adjacent to a non-activating Arg residue so that a Arg-Pro or Pro-Arg dipeptide is created. Complexes of two or more of the analogues, nucleic acids and vectors encoding them may be used alone or in conjunction with each other, for the prevention or treatment of active Factor VIII:C deficiency in a mammal. The analogues may be used as immunogens to raise antibodies, and in the treatment of haemophilias, by improvement of haemostasis. The analogues are resistant to proteolytic cleavage and display increased plasma half-life. They may be administered at lower dosages and by different modes of administration

XX
SQ Sequence 2351 AA;

Query Match 100.0%; Score 12410; DB 2; Length 2351;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2350; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MQIELSTCFELCLRFCSATRYIYGAVLSWDYMQSDLGELPVDARPPRPVKSPFN 60
DB 1 MQIELSTCFELCLRFCSATRYIYGAVLSWDYMQSDLGELPVDARPPRPVKSPFN 60

QY 61 TSVVYKTLFVEFTDHLFNIAKPRPPWMLGPTIQAEVYDVTVITLKNASHPVSLHAV 120
DB 61 TSVVYKTLFVEFTDHLFNIAKPRPPWMLGPTIQAEVYDVTVITLKNASHPVSLHAV 120

QY 121 GVSWKASEGAEYDDQTSQREKDDKVPFGSGHTYVQVLKENGPMASDPLCLTYSLSH 180
DB 121 GVSWKASEGAEYDDQTSQREKDDKVPFGSGHTYVQVLKENGPMASDPLCLTYSLSH 180

QY 181 VDLVKDLNSGLIGALLVCREGSLAKEKTOILHFKIILFPAYFDEGKSHSEFTKSLMQDRD 240
DB 181 VDLVKDLNSGLIGALLVCREGSLAKEKTOILHFKIILFPAYFDEGKSHSEFTKSLMQDRD 240

QY 241 AASARAPKMTYNGYVNRSLPGILGCHRSVYVHVIQMGTTPEVHSIFLFGHTFLVRNH 300
DB 241 AASARAPKMTYNGYVNRSLPGILGCHRSVYVHVIQMGTTPEVHSIFLFGHTFLVRNH 300

QY 301 ROASLEISPTITFLTAQTLLMDLGOFLFCHISSHQHDGMEAYKVVDSCPEPQLRMKNE 360
DB 301 ROASLEISPTITFLTAQTLLMDLGOFLFCHISSHQHDGMEAYKVVDSCPEPQLRMKNE 360

QY 361 EADYDDDLTDSMDVTRFDDNSPFIQIRSVAKKHPTWVHYTAAEEDWDYAPLVLA 420
DB 361 EADYDDDLTDSMDVTRFDDNSPFIQIRSVAKKHPTWVHYTAAEEDWDYAPLVLA 420

QY 421 PDORSYKSOYLNNNGPQIRGRYKVRFMAYTDETEKTRAIQHESGILGPLLYGEVDTL 480
DB 421 PDORSYKSOYLNNNGPQIRGRYKVRFMAYTDETEKTRAIQHESGILGPLLYGEVDTL 480

QY 481 LIIFKNQASRPYNIYPHGITDVRPLYSRRLPKGVKHLKDFPLPGHIFKYKWTVTVEDGP 540
DB 481 LIIFKNQASRPYNIYPHGITDVRPLYSRRLPKGVKHLKDFPLPGHIFKYKWTVTVEDGP 540

QY 541 TKSDPRCLTRYYSFVNMRDLASGLIGPLLI CYKESVDORGNQIMSKERNVILSVPE 600
DB 541 TKSDPRCLTRYYSFVNMRDLASGLIGPLLI CYKESVDORGNQIMSKERNVILSVPE 600

QY 601 NBSWYLTENIQRLPNAGVQLEDPEFQASNMHMSINGVYVDSLOISVCLHEVAYWYILS 660
DB 601 NBSWYLTENIQRLPNAGVQLEDPEFQASNMHMSINGVYVDSLOISVCLHEVAYWYILS 660

QY 661 IGAQTDFLSVFSGYTFKHWVYEDTLTLFPFSGETVFMSPNPGMLWILGCHNSDFNRG 720
DB 661 IGAQTDFLSVFSGYTFKHWVYEDTLTLFPFSGETVFMSPNPGMLWILGCHNSDFNRG 720

CC that ultimately converts soluble fibrinogen to insoluble fibrin clot, effecting haemostasis. A deficiency in factor VIII:C is responsible for haemophilia A, which is an X-chromosome-linked inherited bleeding diathesis. Factor VIII:C is activated by plasma proteases, such as thrombin. During activation the mature polypeptide is cleaved to generate heavy and light chain fragments that are further cleaved. The DNA encoding this sequence is mutated, using mutagenic primers, to produce the active Factor VIII:C analogues of the invention (such as AAW11330). The analogues comprise a native Factor VIII:C polypeptide modified at a site adjacent to a non-activating Arg residue so that a Arg-Pro or Pro-Arg dipeptide is created. Complexes of two or more of the analogues, nucleic acids and vectors encoding them may be used alone or in conjunction with each other, for the prevention or treatment of active Factor VIII:C deficiency in a mammal. The analogues may be used as immunogens to raise antibodies, and in the treatment of haemophilias, by improvement of haemostasis. The analogues are resistant to proteolytic cleavage and display increased plasma half-life. They may be administered at lower dosages and by different modes of administration

XX
SQ Sequence 2351 AA;

Query Match 100.0%; Score 12410; DB 2; Length 2351;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2350; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MQIELSTCFELCLRFCSATRYIYGAVLSWDYMQSDLGELPVDARPPRPVKSPFN 60
DB 1 MQIELSTCFELCLRFCSATRYIYGAVLSWDYMQSDLGELPVDARPPRPVKSPFN 60

QY 61 TSVVYKTLFVEFTDHLFNIAKPRPPWMLGPTIQAEVYDVTVITLKNASHPVSLHAV 120
DB 61 TSVVYKTLFVEFTDHLFNIAKPRPPWMLGPTIQAEVYDVTVITLKNASHPVSLHAV 120

QY 121 GVSWKASEGAEYDDQTSQREKDDKVPFGSGHTYVQVLKENGPMASDPLCLTYSLSH 180
DB 121 GVSWKASEGAEYDDQTSQREKDDKVPFGSGHTYVQVLKENGPMASDPLCLTYSLSH 180

QY 181 VDLVKDLNSGLIGALLVCREGSLAKEKTOILHFKIILFPAYFDEGKSHSEFTKSLMQDRD 240
DB 181 VDLVKDLNSGLIGALLVCREGSLAKEKTOILHFKIILFPAYFDEGKSHSEFTKSLMQDRD 240

QY 241 AASARAPKMTYNGYVNRSLPGILGCHRSVYVHVIQMGTTPEVHSIFLFGHTFLVRNH 300
DB 241 AASARAPKMTYNGYVNRSLPGILGCHRSVYVHVIQMGTTPEVHSIFLFGHTFLVRNH 300

QY 301 ROASLEISPTITFLTAQTLLMDLGOFLFCHISSHQHDGMEAYKVVDSCPEPQLRMKNE 360
DB 301 ROASLEISPTITFLTAQTLLMDLGOFLFCHISSHQHDGMEAYKVVDSCPEPQLRMKNE 360

QY 361 EADYDDDLTDSMDVTRFDDNSPFIQIRSVAKKHPTWVHYTAAEEDWDYAPLVLA 420
DB 361 EADYDDDLTDSMDVTRFDDNSPFIQIRSVAKKHPTWVHYTAAEEDWDYAPLVLA 420

QY 421 PDORSYKSOYLNNNGPQIRGRYKVRFMAYTDETEKTRAIQHESGILGPLLYGEVDTL 480
DB 421 PDORSYKSOYLNNNGPQIRGRYKVRFMAYTDETEKTRAIQHESGILGPLLYGEVDTL 480

QY 481 LIIFKNQASRPYNIYPHGITDVRPLYSRRLPKGVKHLKDFPLPGHIFKYKWTVTVEDGP 540
DB 481 LIIFKNQASRPYNIYPHGITDVRPLYSRRLPKGVKHLKDFPLPGHIFKYKWTVTVEDGP 540

QY 541 TKSDPRCLTRYYSFVNMRDLASGLIGPLLI CYKESVDORGNQIMSKERNVILSVPE 600
DB 541 TKSDPRCLTRYYSFVNMRDLASGLIGPLLI CYKESVDORGNQIMSKERNVILSVPE 600

QY 601 NBSWYLTENIQRLPNAGVQLEDPEFQASNMHMSINGVYVDSLOISVCLHEVAYWYILS 660
DB 601 NBSWYLTENIQRLPNAGVQLEDPEFQASNMHMSINGVYVDSLOISVCLHEVAYWYILS 660

QY 661 IGAQTDFLSVFSGYTFKHWVYEDTLTLFPFSGETVFMSPNPGMLWILGCHNSDFNRG 720
DB 661 IGAQTDFLSVFSGYTFKHWVYEDTLTLFPFSGETVFMSPNPGMLWILGCHNSDFNRG 720

721 MTALLKVSCKNTGDDYEDSYEDISAYLSKXNAIEPRFSQNSRHSTPQKQFNATTTI 780
721 MTALLKVSCKNTGDDYEDSYEDISAYLSKXNAIEPRFSQNSRHSTPQKQFNATTTI 780
781 PENDIEKTDWFAHRTMPKIQNVSSDLMQLQSPTPHGLSLSDLOEAKYETFSDDPS 840
781 PENDIEKTDWFAHRTMPKIQNVSSDLMQLQSPTPHGLSLSDLOEAKYETFSDDPS 840
841 PGALDSNNISSEMTHFAPQLHISGDMVFTPESGLOLRNEKLGTAATELKKLDFKVSST 900
841 PGALDSNNISSEMTHFAPQLHISGDMVFTPESGLOLRNEKLGTAATELKKLDFKVSST 900
901 SNNLISITPSNLAAGTNDTSSIGPPSNVHYDQLODTTLPGKKSSPLTESGGPLSLEE 960
901 SNNLISITPSNLAAGTNDTSSIGPPSNVHYDQLODTTLPGKKSSPLTESGGPLSLEE 960
961 NNDKLLLESGLMNSQESSWGKNVSTESGRFLPKGRAGHPALLTKDNALFKVSISSLKTN 1020
961 NNDKLLLESGLMNSQESSWGKNVSTESGRFLPKGRAGHPALLTKDNALFKVSISSLKTN 1020
1021 KTSNNSATNRKTHIDGSPILLIENSPVMQNTLESDETEKKVTPLIHDMMLDKNATLRL 1080
1021 KTSNNSATNRKTHIDGSPILLIENSPVMQNTLESDETEKKVTPLIHDMMLDKNATLRL 1080
1081 NMSNKTTSKXNMWVQKQCGPFPDPAQNPDMSFFKMLFLPESARWQRTGKNSLNSG 1140
1081 NMSNKTTSKXNMWVQKQCGPFPDPAQNPDMSFFKMLFLPESARWQRTGKNSLNSG 1140
1141 QGSPKQVLSIGPEKSVBEGONFLSKNVVKGGEFTKQVGLKEMVFPSSRNLFUTNLDN 1200
1141 QGSPKQVLSIGPEKSVBEGONFLSKNVVKGGEFTKQVGLKEMVFPSSRNLFUTNLDN 1200
1201 LHENNTNHOEKKIOBEIEKKETLIGENVVLPOIHTVTGTHQPMKNLFLLSTRQNVGSGYD 1260
1201 LHENNTNHOEKKIOBEIEKKETLIGENVVLPOIHTVTGTHQPMKNLFLLSTRQNVGSGYD 1260
1261 GAYAVLQDFRSLNDSTNRKKGTAHFKSKGGEENLEGIGNQTKQIVEKYACTTRISPT 1320
1261 GAYAVLQDFRSLNDSTNRKKGTAHFKSKGGEENLEGIGNQTKQIVEKYACTTRISPT 1320
1321 SQQNFVQSRKXALKQFRLPLEETELEKRIIVDDTSTQMSKXKHLPSTLTQIDYNEKE 1380
1321 SQQNFVQSRKXALKQFRLPLEETELEKRIIVDDTSTQMSKXKHLPSTLTQIDYNEKE 1380
1381 KGAIQTQSLSDCLTRSHSIPQANRSLPIAKVSSPSPRIPIYLRVLPFODNSSHLPASYS 1440
1381 KGAIQTQSLSDCLTRSHSIPQANRSLPIAKVSSPSPRIPIYLRVLPFODNSSHLPASYS 1440
1441 RKKGSGVQSSSHFLOGAKKNNLSLAILTLEMTGDQREVGLGTSAINSVTYKKVENTVLP 1500
1441 RKKGSGVQSSSHFLOGAKKNNLSLAILTLEMTGDQREVGLGTSAINSVTYKKVENTVLP 1500
1501 KPDLPKTSKVELLPKVHIYQKDLFPPTETSNQSPGHLDLVEGSLGCTGCAIKWNEANRP 1560
1501 KPDLPKTSKVELLPKVHIYQKDLFPPTETSNQSPGHLDLVEGSLGCTGCAIKWNEANRP 1560
1561 GXVFFLRVATSSAKTPSKLLDPLANDNHYGTQIPKEWKSQKSEKPAFKKOTILSL 1620
1561 GXVFFLRVATSSAKTPSKLLDPLANDNHYGTQIPKEWKSQKSEKPAFKKOTILSL 1620
1621 NACESNPAIAAINEGNQKPEIEVTWAKQGRTERLCSQNPVLRKHORETRITLQSDQBE 1680
1621 NACESNPAIAAINEGNQKPEIEVTWAKQGRTERLCSQNPVLRKHORETRITLQSDQBE 1680
1681 IDYDDTISVEMKXEDFDIYDEDENQSPRSQKTRHYFFIAAVERLWDYGMSSSPHYLRNR 1740
1681 IDYDDTISVEMKXEDFDIYDEDENQSPRSQKTRHYFFIAAVERLWDYGMSSSPHYLRNR 1740
1741 AQSGSVQPKKVVQFQFTDGSFTQPLVYRGLNHLGLLGPYIRAEVEDNIMVTRNQASR 1800
1741 AQSGSVQPKKVVQFQFTDGSFTQPLVYRGLNHLGLLGPYIRAEVEDNIMVTRNQASR 1800
1801 PYSFYSSLISYEEDQRCQAEPRKNFVKPNETKTYFWKVQHMAPTKDBDFCKAWAYFSDV 1860

1801 PYSFYSSLISYEEDQRCQAEPRKNFVKPNETKTYFWKVQHMAPTKDBDFCKAWAYFSDV 1860
1861 DLEKDVHSGLIGLIVCHTNTLNPAGROVTVQEFALFFTFIDETKSWYFTEWNERCRA 1920
1861 DLEKDVHSGLIGLIVCHTNTLNPAGROVTVQEFALFFTFIDETKSWYFTEWNERCRA 1920
1921 PCNIQMEDPTFKENYRFAHNGYIMDTLPGLVNAQDQRIKWYLLSMGSENENIHSIFSGH 1980
1921 PCNIQMEDPTFKENYRFAHNGYIMDTLPGLVNAQDQRIKWYLLSMGSENENIHSIFSGH 1980
1981 VFTVRKKEBYKVALYNLYPGVFETVEMLPKAGIWRVECLIGELHAGMSTLFLVYSNKC 2040
1981 VFTVRKKEBYKVALYNLYPGVFETVEMLPKAGIWRVECLIGELHAGMSTLFLVYSNKC 2040
2041 QTPILGASGHIRDFQITASGOYGOWAPKLARLHVSGINAWSTKEPPSWIKVDLLAPMI I 2100
2041 QTPILGASGHIRDFQITASGOYGOWAPKLARLHVSGINAWSTKEPPSWIKVDLLAPMI I 2100
2101 HGIKTQARQKFSSLYISQFIIMVSLDGKKWQTVYRGNSTGTLMVFFGNVDSGSIKKNIFN 2160
2101 HGIKTQARQKFSSLYISQFIIMVSLDGKKWQTVYRGNSTGTLMVFFGNVDSGSIKKNIFN 2160
2161 PPIIARIYIRLHPHYSTIRLMELMCGDLNSCMLPGMESKALSDAQITASSYFTNWEA 2220
2161 PPIIARIYIRLHPHYSTIRLMELMCGDLNSCMLPGMESKALSDAQITASSYFTNWEA 2220
2221 TWSFSKARLHLOGRSNAWRPOVNNPKEWLQVDFOKTMKVGTQVTKGKSLTSMYVKEFL 2280
2221 TWSFSKARLHLOGRSNAWRPOVNNPKEWLQVDFOKTMKVGTQVTKGKSLTSMYVKEFL 2280
2281 ISSSQDGHQWTLFPQNGKVKVFGQNQDSFTPVVNSLDPPLLTRYLRIHPQSWVHQIALRM 2340
2281 ISSSQDGHQWTLFPQNGKVKVFGQNQDSFTPVVNSLDPPLLTRYLRIHPQSWVHQIALRM 2340
2341 EVLGCRAQDLY 2351
2341 EVLGCRAQDLY 2351
RESULT 3
AAW46245
ID AAW46245 standard; protein; 2351 AA.
XX
AC AAW46245;
XX
DT 25-MAR-2003 (revised)
DT 06-AUG-1998 (first entry)
XX
Human factor III protein sequence.
XX
Replication defective; recombinant retrovirus; RRV; therapeutic protein;
KW haemophilia; thrombosis; hypercoagulable disorder; liver disease; human;
KW hepatic; thalassemia; phenylketonuria; Lesch-Nyhan syndrome; diabetes;
KW cystic fibrosis; Duchenne's Muscular Dystrophy; hypercholesterolemia;
KW hypopituitarism; adenine deaminase deficiency; HIV infection; anaemia;
KW Guacher's syndrome; high blood pressure; Alzheimer's disease, factor III;
KW autoimmune; inflammatory disease.
XX
Homo sapiens.
XX
WO9800541-A2.
XX
08-JAN-1998.
XX
02-JUL-1997; 97WO-US011784.
XX
03-JUL-1996; 96US-00645601.
PR 13-AUG-1996; 96US-00696381.
PR 04-JUN-1997; 97US-00869309.
XX
(CHIR) CHIRON CORP.
XX

PI Jolly DJ, Barber JR, Chang SMW, Respass JG, Allen JR, Boder M;
 PI Chong K, De La Vega D, Depolo NJ, Hsu DC, Ibanez CE;
 PI Mittelstaedt DM, Prussak CE, Greengard J, Lee R;
 XX WPI; 1998-086966/08.
 DR N-PSDB; AAV19580.

XX New replication defective recombinant retro-viruses - which can be
 PT administered to provide long term systemic expression of therapeutic
 PT protein in blood, useful in, e.g. treating hyper-coagulable disorders.
 XX
 XX Example 28; Page 203-210; 272pp; English.

XX This is the human factor III sequence. The encoding DNA is used in the
 CC construction of recombinant retroviral vectors expressing human factor
 CC VIII. The specification provides the preparation of replication defective
 CC recombinant retrovirus (RRV) expressing a therapeutic protein. The RRV
 CC preparation is resistant to degradation by human complement and is
 CC capable of inducing long term systemic expression of the therapeutic
 CC protein when administered intravenously to a human. The long term
 CC systemic expression results in a measurable level of the therapeutic
 CC protein being produced in the blood of the human for a period of at least
 CC 30 days after the administration of the RRV vector preparation. RRV's can
 CC be used for in vivo delivery of therapeutic protein to treat, e.g.
 CC haemophilia A, haemophilia B, thrombosis, hypercoagulable disorders,
 CC liver diseases such as hepatitis, disorders such as thalassemia,
 CC phenylketonuria, Lesch-Nyhan syndrome, severe combined immunodeficiency
 CC (SCID), cystic fibrosis, Duchenne's Muscular Dystrophy, inherited
 CC emphysema, familial hypercholesterolemia, diabetes, hypopituitarism,
 CC ageneine deaminase deficiency, alapha-antitrypsin deficiency, Guacher's
 CC syndrome, anaemia, infections such as HIV infection, high blood pressure,
 CC Alzheimer's disease, autoimmune or inflammatory disease or graft versus
 CC host disease. RRV's are capable of surviving inactivation in human serum
 CC thereby allowing efficient gene transfer over prolonged periods of time.
 XX (Updated on 25-MAR-2003 to correct PI field.)

SQ Sequence 2351 AA;

Query Match 100.0%; Score 12410; DB 2; Length 2351;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 2350; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MOELSTCFCLLRFCFSATRRYVILGAVELSWDQSLGELPVDARPPRPVPSFFPN 60
 DB 1 MOELSTCFCLLRFCFSATRRYVILGAVELSWDQSLGELPVDARPPRPVPSFFPN 60

QY 61 TSVVYKTLFVEFTDHLFNIAKRP2PMGLLQFTIOAEVYDVTWITLKNMASHPVSLHAV 120
 DB 61 TSVVYKTLFVEFTDHLFNIAKRP2PMGLLQFTIOAEVYDVTWITLKNMASHPVSLHAV 120

QY 121 GVSYYKASGAYDDOTSOREKEDDKVFGSGHTYVQVLKENGPMASDPLCLTYSLSH 180
 DB 121 GVSYYKASGAYDDOTSOREKEDDKVFGSGHTYVQVLKENGPMASDPLCLTYSLSH 180

QY 181 VDLVLDVNSGLIGALLVCRGSLAKEKTQLHKFILLFAVDFEGKSWHSETKNSLMQDRD 240
 DB 181 VDLVLDVNSGLIGALLVCRGSLAKEKTQLHKFILLFAVDFEGKSWHSETKNSLMQDRD 240

QY 241 AASARAWPQHTVNGVYVNSRLPGLIGCHRSKVYVHVGMTTPEVHSIFLEGHTFLVRNH 300
 DB 241 AASARAWPQHTVNGVYVNSRLPGLIGCHRSKVYVHVGMTTPEVHSIFLEGHTFLVRNH 300

QY 301 ROASLEISITITLTAQTLMDLQGFLLSCHISSHQHDGMEAYVKVDSCEEPQLRMKNE 360
 DB 301 ROASLEISITITLTAQTLMDLQGFLLSCHISSHQHDGMEAYVKVDSCEEPQLRMKNE 360

QY 361 EAEDYDLDLTDSEMDVVRDSDNSPSFIOIRSVAKKHPTWVHYIAAEBEDMDYAPLVLA 420
 DB 361 EAEDYDLDLTDSEMDVVRDSDNSPSFIOIRSVAKKHPTWVHYIAAEBEDMDYAPLVLA 420

QY 421 PDRSYKSYQLNKGPRIOIRKYYKVFPMAYDTDTFTKTREAIQHSIGILGPLLYGEVGDTL 480
 DB 421 PDRSYKSYQLNKGPRIOIRKYYKVFPMAYDTDTFTKTREAIQHSIGILGPLLYGEVGDTL 480

QY 481 LLIIFKQASRPYNIYPHGIITDVRLPYSRRLPKGVKHLKDPFLLPGEIFKFKWTVVBGP 540
 DB 481 LLIIFKQASRPYNIYPHGIITDVRLPYSRRLPKGVKHLKDPFLLPGEIFKFKWTVVBGP 540

QY 541 TKSDPRCLTRYSSFVNMRDLASGLIGPLLIICYKESVDQRGNQIMSDKRNVLFSVFDE 600
 DB 541 TKSDPRCLTRYSSFVNMRDLASGLIGPLLIICYKESVDQRGNQIMSDKRNVLFSVFDE 600

QY 601 NRSWYLNTQRFELNPNAGVQLEDPEQASNMHSINGYVFDLSQLSVCLHEVAYWILS 660
 DB 601 NRSWYLNTQRFELNPNAGVQLEDPEQASNMHSINGYVFDLSQLSVCLHEVAYWILS 660

QY 661 IGAQTDLSLFFSYGTYFKHWYEDTTLTPPFGSETVFMGMENPGLWTLGCHNSDFNRNG 720
 DB 661 IGAQTDLSLFFSYGTYFKHWYEDTTLTPPFGSETVFMGMENPGLWTLGCHNSDFNRNG 720

QY 721 MTALLKVSSCDKNTGDIYEDSYEDISAYLLSKNNAIEPRSFQNSRHPSTKQFNATTI 780
 DB 721 MTALLKVSSCDKNTGDIYEDSYEDISAYLLSKNNAIEPRSFQNSRHPSTKQFNATTI 780

QY 781 PENDIEKTDWFAHRTMPKIQNVSSDLMLLRQSPTPHGLSLSDLOEAKYETFSDDPS 840
 DB 781 PENDIEKTDWFAHRTMPKIQNVSSDLMLLRQSPTPHGLSLSDLOEAKYETFSDDPS 840

QY 841 PGALDSNNSISEMTHFRPOLHHSQDMVTTPBSGLQLRLNEKLGTATAELKLDKFKVSS 900
 DB 841 PGALDSNNSISEMTHFRPOLHHSQDMVTTPBSGLQLRLNEKLGTATAELKLDKFKVSS 900

QY 901 SNNLITIPSDNLAAAGTNTSSLGPPNPMVHYDSOLDTTLFGKSSPLTESGGPLSLSEE 960
 DB 901 SNNLITIPSDNLAAAGTNTSSLGPPNPMVHYDSOLDTTLFGKSSPLTESGGPLSLSEE 960

QY 961 NNDKLLLESGLMNSQESSWGKNVASTESGRVFKGRAGPALLTKDNALFKVSIALLKTN 1020
 DB 961 NNDKLLLESGLMNSQESSWGKNVASTESGRVFKGRAGPALLTKDNALFKVSIALLKTN 1020

QY 1021 KTSNKSATNRKTHIDGPSLLIENSFVWQNILEDSTEFKKVTPLIHRLMDKNATLRL 1080
 DB 1021 KTSNKSATNRKTHIDGPSLLIENSFVWQNILEDSTEFKKVTPLIHRLMDKNATLRL 1080

QY 1081 NHMKNKTTSSKNMVMVQKKGPIPPDAQNQDMDFPKMLFIPESARWQORTHGKNSLNG 1140
 DB 1081 NHMKNKTTSSKNMVMVQKKGPIPPDAQNQDMDFPKMLFIPESARWQORTHGKNSLNG 1140

QY 1141 QGSPKQVLSGLPEKSVGQNFLEKKNVVGKGFETKDVGLKENAVFPSSRNLFITNLDN 1200
 DB 1141 QGSPKQVLSGLPEKSVGQNFLEKKNVVGKGFETKDVGLKENAVFPSSRNLFITNLDN 1200

QY 1201 LHENNTNQEKKIQEIEIEKKETLIQENVVLPQIHVTOTKPMKNLFLLSRQNVGSDY 1260
 DB 1201 LHENNTNQEKKIQEIEIEKKETLIQENVVLPQIHVTOTKPMKNLFLLSRQNVGSDY 1260

QY 1261 GAYAVLPQDPSLNDSTNRKHTAHFSGKEENLEGNGQTKQIVKCYACTTISNT 1320
 DB 1261 GAYAVLPQDPSLNDSTNRKHTAHFSGKEENLEGNGQTKQIVKCYACTTISNT 1320

QY 1321 SOONFVTOQRKALQFRLPLEETELEKRIIVDDTSTQWSKMKHLTPSTLTQIDYNEKE 1380
 DB 1321 SOONFVTOQRKALQFRLPLEETELEKRIIVDDTSTQWSKMKHLTPSTLTQIDYNEKE 1380

QY 1381 KGAIQTQSPLSDCLTRSHSIPQANRSPLPIAKVSSFPFSIRPIYLRVLFDQNSSHLPAA 1440
 DB 1381 KGAIQTQSPLSDCLTRSHSIPQANRSPLPIAKVSSFPFSIRPIYLRVLFDQNSSHLPAA 1440

QY 1441 RKQSGVQESSHFLGAKAKNNLSAILT-EMTGDQREVGSGLTSATNSVTKKVENTVLP 1500
 DB 1441 RKQSGVQESSHFLGAKAKNNLSAILT-EMTGDQREVGSGLTSATNSVTKKVENTVLP 1500

QY 1501 KPDLPKTSKVBELLPKVHIYQKDLPTTSTNGSPGHLDLVEGSLLOGTEGALKWNEANRP 1560
 DB 1501 KPDLPKTSKVBELLPKVHIYQKDLPTTSTNGSPGHLDLVEGSLLOGTEGALKWNEANRP 1560

Qy 1561 GKVPFLVATSSAKTBSKLLDPLAWDNHGYGTQPKPEWKSQBSPEKTAFAKKXDTILSL 1620
 Db |||||
 Qy 1561 GKVPFLVATSSAKTBSKLLDPLAWDNHGYGTQPKPEWKSQBSPEKTAFAKKXDTILSL 1620
 Db |||||
 Qy 1621 NACESNHAIAAINBGNKPELEVWAKQGRTERLCNQPPVLYKHPQREITRTTLQSDQEE 1680
 Db |||||
 Qy 1621 NACESNHAIAAINBGNKPELEVWAKQGRTERLCNQPPVLYKHPQREITRTTLQSDQEE 1680
 Db |||||
 Qy 1661 IDYDDTTSVENMKKEDFDIYDEENQSPRSFQKTRHYZAAVERLDWYGMSSSHVLRNR 1740
 Db |||||
 Qy 1681 IDYDDTTSVENMKKEDFDIYDEENQSPRSFQKTRHYFAAVERLDWYGMSSSHVLRNR 1740
 Db |||||
 Qy 1741 AQSGSVPOFKKVVPOEFTDGGFTQPLYGELNEHLGLGPYIRAEVDENIIVTFENQASR 1800
 Db |||||
 Qy 1741 AQSGSVPOFKKVVPOEFTDGGFTQPLYGELNEHLGLGPYIRAEVDENIIVTFENQASR 1800
 Db |||||
 Qy 1801 PYSFYSSLSIYEEDQOQAEPRKNVKNETKYPKWQVQHNAPYKDFDCKAWAYFSDV 1860
 Db |||||
 Qy 1801 PYSFYSSLSIYEEDQOQAEPRKNVKNETKYPKWQVQHNAPYKDFDCKAWAYFSDV 1860
 Db |||||
 Qy 1861 DLEKDVHSLIGPLIVCHTNTLNPAHGRQVTVQBEALFTTFDETKSWYFTENMERNCR 1920
 Db |||||
 Qy 1861 DLEKDVHSLIGPLIVCHTNTLNPAHGRQVTVQBEALFTTFDETKSWYFTENMERNCR 1920
 Db |||||
 Qy 1921 PCNIQWEDPTFKENYRFHAINGYINDTLPLGLVMAQDQIRWYILSMGSGNENIHSIHFSGH 1980
 Db |||||
 Qy 1921 PCNIQWEDPTFKENYRFHAINGYINDTLPLGLVMAQDQIRWYILSMGSGNENIHSIHFSGH 1980
 Db |||||
 Qy 1981 VFTVRKKEEYKMALNLYPGVPETVEMLPKAGIWRVECLIGEHLHAGMSTLFLVYSNKC 2040
 Db |||||
 Qy 1981 VFTVRKKEEYKMALNLYPGVPETVEMLPKAGIWRVECLIGEHLHAGMSTLFLVYSNKC 2040
 Db |||||
 Qy 2041 QPLGWASGHIRDFQTASGQYQWAPKLARLHYSGSINAWSTKEPPSWIKVDLLAPMLI 2100
 Db |||||
 Qy 2041 QPLGWASGHIRDFQTASGQYQWAPKLARLHYSGSINAWSTKEPPSWIKVDLLAPMLI 2100
 Db |||||
 Qy 2101 HGKIQGARQKFSLSYISQFIIMYSLDGKKWQTYRGNSTGTLWFFGNVDSSGIXHNIPN 2160
 Db |||||
 Qy 2101 HGKIQGARQKFSLSYISQFIIMYSLDGKKWQTYRGNSTGTLWFFGNVDSSGIXHNIPN 2160
 Db |||||
 Qy 2161 PPIIARYIRLHPTHYSIRTLRMELMGCDLNSCPLGWESKAISDAQITASSYFTNMEA 2220
 Db |||||
 Qy 2161 PPIIARYIRLHPTHYSIRTLRMELMGCDLNSCPLGWESKAISDAQITASSYFTNMEA 2220
 Db |||||
 Qy 2221 TWSPSKARLHLOGRNAPQVNNPKWLQVDFQKTMKVGTGTVQVKSLLTSMYVKEFL 2280
 Db |||||
 Qy 2221 TWSPSKARLHLOGRNAPQVNNPKWLQVDFQKTMKVGTGTVQVKSLLTSMYVKEFL 2280
 Db |||||
 Qy 2281 ISSSQDGHQWTLFFQNGKVKVFGNQDSFTPVVNSLDPPLLYRLRIHPQSWVHQIALRM 2340
 Db |||||
 Qy 2281 ISSSQDGHQWTLFFQNGKVKVFGNQDSFTPVVNSLDPPLLYRLRIHPQSWVHQIALRM 2340
 Db |||||
 Qy 2341 EVLGCFAQDLY 2351
 Db |||||
 Qy 2341 EVLGCFAQDLY 2351
 Db |||||

RESULT 4

AAW44373

ID AAW44373 standard; protein; 2351 AA.

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

Domain 711..1694
 /note="B-domain"
 WO9800542-A2.
 08-JAN-1998.
 02-JUL-1997; 97WO-US0111785.
 03-JUL-1996; 96US-00645601.
 13-AUG-1996; 96US-00696381.
 04-JUN-1997; 97US-00869309.
 (CHIR) CHIRON CORP.
 Jolly DJ, Barber JR, Chang SM, Respass JG, Allen JR, Boder M;
 Chong K, De La Vega D, Depolo NJ, Hsu DC, Ibanez CE;
 Mittelstaedt DM, Prussak CE, Greengard J;
 WPI; 1998-086967/08.
 N-PSDB; AAV15359.
 New replication defective recombinant retroviruses - which express B
 domain-deleted human factor VIII or human factor IX for the treatment of
 haemophilia.
 Example 28; Page 164-166; 236pp; English.
 This polypeptide comprises human Factor VIII. The invention relates to
 preparations of replication defective recombinant retrovirus (RV)
 expressing a B domain-deleted human factor VIII protein, such as the SQN
 deletion mutant (see AAW4372), where the recombinant RV is capable of
 infecting human cells, is resistant to degradation by human complement
 and is capable of inducing long-term (at least 30 days and up to 6 months
 or longer post-injection) systemic expression of factor VIII when
 administered to a haemophilia A patient
 Sequence 2351 AA;
 Query Match 100.0%; Score 12410; DB 2; Length 2351;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 2350; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 Qy 1 MQIELSTCFFCLLRFPCFSATRRYYLGAVELSWDMQSDLGELPVDARPPRVPKSFPPN 60
 Db 1 MQIELSTCFFCLLRFPCFSATRRYYLGAVELSWDMQSDLGELPVDARPPRVPKSFPPN 60
 Qy 61 TSVVYKKTILFVEPTDHLNIAKPRPPNMGLLGFTTQAEVYDVITLKNMASHPVSLHAV 120
 Db 61 TSVVYKKTILFVEPTDHLNIAKPRPPNMGLLGFTTQAEVYDVITLKNMASHPVSLHAV 120
 Qy 121 GVSVMKASGEGAYDDQTSOREKEDDKVFPGGSHYVWQVLKENGPMASDPLCLTYSYLH 180
 Db 121 GVSVMKASGEGAYDDQTSOREKEDDKVFPGGSHYVWQVLKENGPMASDPLCLTYSYLH 180
 Qy 181 VDLVXDLNSGLIGALLVCREGSLAKEKTQTLHKFILLFAVDEGKSWSETKNSLMQDD 240
 Db 181 VDLVXDLNSGLIGALLVCREGSLAKEKTQTLHKFILLFAVDEGKSWSETKNSLMQDD 240
 Qy 241 AASARAWPKQHTVNGYVNRSLPGLIGCHRKSVYWHVIGMTTPEVHSIFLEGHTFLVRNH 300
 Db 241 AASARAWPKQHTVNGYVNRSLPGLIGCHRKSVYWHVIGMTTPEVHSIFLEGHTFLVRNH 300
 Qy 301 ROASLEISPIITELTAQTLLMDLGQFLLSCHISSHQHDGMEAYVKVDSCPEEPOLMKNE 360
 Db 301 ROASLEISPIITELTAQTLLMDLGQFLLSCHISSHQHDGMEAYVKVDSCPEEPOLMKNE 360
 Qy 361 EADYDDDDTDSMDVVRPDDDNSSPSFIQIRSVAKKHPKTVWHYIAAEEDWDYAPLVIA 420
 Db 361 EADYDDDDTDSMDVVRPDDDNSSPSFIQIRSVAKKHPKTVWHYIAAEEDWDYAPLVIA 420
 Qy 421 PDRSVKSOYLNNGPORIGKVKYRFEMAYTDTETFKTREAIQHOESGILGPLYGEGDTL 480
 Db |||||

Location/Qualifiers

421 PDDRSYKSYLNNGPQRIKGRKXKVRFMAYTDETFKTRTREAQHESGILGPLLYGEVGDITL 480
481 LIIPKQASRPYNTYPHGITDVRPLYSRRIPKGVKHLKDPILPGBIPKYNWTVVZDGP 540
481 LIIPKQASRPYNTYPHGITDVRPLYSRRIPKGVKHLKDPILPGBIPKYNWTVVZDGP 540
541 TKSDPRLCTRYYSFVNMERDASGLIGPLLCYKESVDQRGNQIMSDKRNVLIFSVD 600
541 TKSDPRLCTRYYSFVNMERDASGLIGPLLCYKESVDQRGNQIMSDKRNVLIFSVD 600
601 NRSWLTENIQRLFPNAGVQLEDPEFQASNIMHSINGYVDFS:QLSVCLHEVAYWILS 660
601 NRSWLTENIQRLFPNAGVQLEDPEFQASNIMHSINGYVDFS:QLSVCLHEVAYWILS 660
661 IGAOTDELVSVPFGYTFKHKVAVZD:LTLPFSGETVFMSENPLMTILGCHNSDFNRG 720
661 IGAOTDELVSVPFGYTFKHKVAVZD:LTLPFSGETVFMSENPLMTILGCHNSDFNRG 720
721 MTALLKVSCKNTGDYVEDSVEDISAYLLSKNNAIEPRSPSONSRHPS:TRQKQFNATTI 780
721 MTALLKVSCKNTGDYVEDSVEDISAYLLSKNNAIEPRSPSONSRHPS:TRQKQFNATTI 780
781 PENDIEKTDPAFHRTMPKIPONTSSDMLLROSPTPHGLSISDIOEAKYETPSDDPS 840
781 PENDIEKTDPAFHRTMPKIPONTSSDMLLROSPTPHGLSISDIOEAKYETPSDDPS 840
841 PGADSNLSSEYTHFRPQLHHSQDMVTPESGLQRLNEKLGTTAA:TELKGLDFKVSST 900
841 PGADSNLSSEYTHFRPQLHHSQDMVTPESGLQRLNEKLGTTAA:TELKGLDFKVSST 900
901 SNNLITIPSDNLAAGTNTSLGPPSPVHYDSQD:DTTLFGKSSPLTESGGP:LSLEE 960
901 SNNLITIPSDNLAAGTNTSLGPPSPVHYDSQD:DTTLFGKSSPLTESGGP:LSLEE 960
961 NNDKSLBESGLMNSQSSGNKVNSTESGRLPKGRAGPALLTKDNALFKVSI:SLIKTN 1020
961 NNDKSLBESGLMNSQSSGNKVNSTESGRLPKGRAGPALLTKDNALFKVSI:SLIKTN 1020
1021 KTSNNSATNRKTHIDGSPILLIENS:SVWQNI:LESDETEPKKVTPLIHDRMLDKNATALRL 1080
1021 KTSNNSATNRKTHIDGSPILLIENS:SVWQNI:LESDETEPKKVTPLIHDRMLDKNATALRL 1080
1081 NHMSKNTTSSKNMVMQKKEGPIPPDAQNPDMSFFKMLFLPESARW:IQRTGKNSLNSG 1140
1081 NHMSKNTTSSKNMVMQKKEGPIPPDAQNPDMSFFKMLFLPESARW:IQRTGKNSLNSG 1140
1141 QGFSPLQLVSLGPEKSVQGNFLSEKNVVGKGEFTKDVGLKEMVFPSSRNLFITNLDN 1200
1141 QGFSPLQLVSLGPEKSVQGNFLSEKNVVGKGEFTKDVGLKEMVFPSSRNLFITNLDN 1200
1201 LKENNTHOKEKTOBEIEKKEKTIQENNVLP:QHTVTGKNFMKNLFLLS:TRQNVGSYD 1260
1201 LKENNTHOKEKTOBEIEKKEKTIQENNVLP:QHTVTGKNFMKNLFLLS:TRQNVGSYD 1260
1261 GAYAPVLDQFRSLDSTNPTKHTAHFSKGBEENLEGLGNQTKQIVEKYACTTRIS:PNPT 1320
1261 GAYAPVLDQFRSLDSTNPTKHTAHFSKGBEENLEGLGNQTKQIVEKYACTTRIS:PNPT 1320
1321 SQNFVTSKRAKQFRPLBETHELKRI:IVDDTSTQWSKNMKH:TPSTLIDIDNEKE 1380
1321 SQNFVTSKRAKQFRPLBETHELKRI:IVDDTSTQWSKNMKH:TPSTLIDIDNEKE 1380
1381 KGAITQSPSDCTTRSHSIPQANRSP:PIAKVSGSPSIRIY:LTRVLFDQNSSHLPAASY 1440
1381 KGAITQSPSDCTTRSHSIPQANRSP:PIAKVSGSPSIRIY:LTRVLFDQNSSHLPAASY 1440
1441 RKKDSGVQESSHFLQGAKNKLSLA:ITLEMTGDQREVGLGTSATNSVYKXVENTVLP 1500
1441 RKKDSGVQESSHFLQGAKNKLSLA:ITLEMTGDQREVGLGTSATNSVYKXVENTVLP 1500
1501 KPDLPTSGVKEPLLKPVHIIYQKOLFPTETNGSPGHLDLVEGILLQGTESAIKWNEANRP 1560
1501 KPDLPTSGVKEPLLKPVHIIYQKOLFPTETNGSPGHLDLVEGILLQGTESAIKWNEANRP 1560

1561 GKVPFLRVATESSAKTPSKLLDPLAWNHYGTQIPKEWKSQBSKSPKTAFFKKOTILSL 1620
1561 GKVPFLRVATESSAKTPSKLLDPLAWNHYGTQIPKEWKSQBSKSPKTAFFKKOTILSL 1620
1621 NACRSNHAIAAINEGQNKPEIETWAKOGTEBLCSQNPVLPKHOREITRTTLOSQOEE 1680
1621 NACRSNHAIAAINEGQNKPEIETWAKOGTEBLCSQNPVLPKHOREITRTTLOSQOEE 1680
1681 IDYDDTISVMKKEDFDIYDEDEENQSPRSQKTRHYFIAAVERLWDYGMSSSHVLNR 1740
1681 IDYDDTISVMKKEDFDIYDEDEENQSPRSQKTRHYFIAAVERLWDYGMSSSHVLNR 1740
1741 AQSGSVQFKKXVQFEFTDGSFTQPLVRGEINHEHLLGPGVIRAEVDNIMVTRNOASR 1800
1741 AQSGSVQFKKXVQFEFTDGSFTQPLVRGEINHEHLLGPGVIRAEVDNIMVTRNOASR 1800
1801 PYSYSSLSIYSEEDQROGAPRKNFVKPNETKTVFKVQHHVAPTKEFDCAKWAYFSDV 1860
1801 PYSYSSLSIYSEEDQROGAPRKNFVKPNETKTVFKVQHHVAPTKEFDCAKWAYFSDV 1860
1861 DLEKDVHSGILGPLLVCHTNTLNPAHGRQVTVQBFALFFTFIDETKSWYFTENMERNCR 1920
1861 DLEKDVHSGILGPLLVCHTNTLNPAHGRQVTVQBFALFFTFIDETKSWYFTENMERNCR 1920
1921 PCNIQMEDPTPKENYRPHAINGYIMDTLPGLVMAQDQIRWYLLSMGSNENIHSIHFSGH 1980
1921 PCNIQMEDPTPKENYRPHAINGYIMDTLPGLVMAQDQIRWYLLSMGSNENIHSIHFSGH 1980
1981 VFTVRKKEEYKMALYNLYPGVFETVEMLPKAGIWRVECLIGEHLHAGMSTLELVYSNKC 2040
1981 VFTVRKKEEYKMALYNLYPGVFETVEMLPKAGIWRVECLIGEHLHAGMSTLELVYSNKC 2040
2041 QTPLGMSGHRDFOITASGOYQWAPKLABLHYSGSINAWSTKEPFSWIKVDLLAPMII 2100
2041 QTPLGMSGHRDFOITASGOYQWAPKLABLHYSGSINAWSTKEPFSWIKVDLLAPMII 2100
2101 EGIKTQGARQKPSLIYISQFIIMYSLDGKKQWTVRGNSTGTLMVFFGNVDSGIKHENIFN 2160
2101 EGIKTQGARQKPSLIYISQFIIMYSLDGKKQWTVRGNSTGTLMVFFGNVDSGIKHENIFN 2160
2161 PPIIARYTRLHPHTHSIESTIRMEGLMGCDLNSCMLPGLWESKAI:SDAQITASSYTNMFA 2220
2161 PPIIARYTRLHPHTHSIESTIRMEGLMGCDLNSCMLPGLWESKAI:SDAQITASSYTNMFA 2220
2221 TWSPSKARLHLQGRSNARPOVNNPKWLQVDFOKTMKVTVTTQGVKSLLTSMYVKEFL 2280
2221 TWSPSKARLHLQGRSNARPOVNNPKWLQVDFOKTMKVTVTTQGVKSLLTSMYVKEFL 2280
2281 ISSSQDGHQWTLFFQNGKVKVQGNQDSFTPVVNSGLDPPLLTRYLRIHPQSVWHQIALRM 2340
2281 ISSSQDGHQWTLFFQNGKVKVQGNQDSFTPVVNSGLDPPLLTRYLRIHPQSVWHQIALRM 2340
2341 EVLGCEAQDLY 2351
2341 EVLGCEAQDLY 2351

RESULT 5

AA52537
ID AA52537 standard; protein; 2351 AA.

XX AA52537;

XX AC

XX XX

XX 28-FEB-2000 (first entry)

XX Human full-length factor VIII.

DE

XX Factor VIII; haemophilia; proteolysis; heavy chain; light chain;

KW secretion; von Willebrand Factor; vWF; C2 domain; intron 22; inversion;

KW non-functional; bleeding disorder; coagulation; treatment.

XX

OS Homo sapiens.

XX Key Location/Qualifiers
 FH Peptide 1..19 "Signal peptide"
 FT /note= "1708"
 FT Protein /note= "Maximum length human factor VIII heavy chain"
 FT Protein 1709..2351
 FT /note= "Human factor VIII light chain"
 XX WO959622-A1.
 FN
 XX 25-NOV-1999.
 XX
 XX 17-MAY-1999; 99WO-US010872.
 XX
 XX 20-MAY-1998; 98US-00082000.
 XX
 XX (GETH) GENENTECH INC.
 XX Vohar GA;
 XX
 XX MPI: 2000-053195/04.
 DR N-PSDB; AA238604.
 XX
 XX New method for treating hemophilia A.
 PT
 XX
 XX Disclosure; Fig 9; 50pp; English.
 XX
 CC This sequence represents the full-length human factor VIII. In this form,
 CC the protein is 300 kD in size with the domain structure A1-A2-B-A3-C1-C2.
 CC However, prior to secretion, this 300 kD protein is proteolyzed into a
 CC heavy chain (A1-A2-B, with continued proteolysis within the B domain
 CC resulting in molecules of varying length) and a light chain (A3-C1-C2)
 CC that remains non-covalently attached to the heavy chain. Upon secretion,
 CC factor VIII is rapidly cleared from the circulation unless it is bound by
 CC the plasma protein von Willebrand Factor (vWF). Factor VIII binds to vWF
 CC through the light chain, with known binding sites at the N-terminus and
 CC within the C2 domain. Haemophilia A is frequently caused by an intron 22
 CC inversion in the factor VIII gene, which separates the gene into two
 CC sections, exons 1-22 becoming inverted and localised telomeric to the
 CC original site, while exons 23-26 remain isolated at the original site.
 CC Exons 23-26 a portion of the C1 and all of the C2 domains, without which
 CC factor VIII is non-functional. The factor VIII gene product of
 CC individuals with this mutation thus comprises domains A1-A2-B-A3 plus a
 CC fragment of C1, which on proteolysis is non-functional, resulting in a
 CC bleeding disorder. The invention relates to a novel method for for
 CC treating haemophilia A in a mammal able to produce the factor VIII heavy
 CC chain. The method comprises administering to the mammal either an
 CC effective amount of factor VIII light chain, or a gene encoding it, and
 CC may be useful for treating patients such as those whose haemophilia A is
 CC caused by intron 22 inversion. The recombinant factor VIII products of
 CC this invention are derived from well-characterised starting materials
 CC which therefore reduces the risk of pathogenic infection which was
 CC previously a problem when using donated plasma. Furthermore, the
 CC invention provides a more economic and potentially more effective
 CC treatment for haemophilia. There is also a need of providing factor VIII
 CC activity to patients who produce or are at risk of producing antibodies
 CC against full-length factor VIII
 XX
 SQ Sequence 2351 AA;
 Query Match 100.0%; Score 12410; DB 3; Length 2351;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 2350; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 1 MOIELSTCFELCLLRFCSATRYVYLGAVELSDWYSDMOSDLGELPVDARRPPRPVKSPFPN 60
 DB 1 XQIELSTCFELCLLRFCSATRYVYLGAVELSDWYSDMOSDLGELPVDARRPPRPVKSPFPN 60
 QY 61 TSVVYKXTLVEFTDHLFNIAPKRPVWMLGPTIQAEVYDVTVITLKNWASHPVSLHAV 120
 DB 61 TSVVYKXTLVEFTDHLFNIAPKRPVWMLGPTIQAEVYDVTVITLKNWASHPVSLHAV 120

QY 121 GVSYWKASGAEAYDDQTSQREKEDKVPFGSGSHYVWQVLKENGPMASDPLCLTYSYLSH 180
 DB 121 GVSYWKASGAEAYDDQTSQREKEDKVPFGSGSHYVWQVLKENGPMASDPLCLTYSYLSH 180
 QY 181 VDLVKDLSNGSLIGALLVCRGSLAKEXTQTLHKETILLFAVPDEGKSWHSETKNSLMDORD 240
 DB 181 VDLVKDLSNGSLIGALLVCRGSLAKEXTQTLHKETILLFAVPDEGKSWHSETKNSLMDORD 240
 QY 241 AASARAWPKMTVANGYVNRSLPGLTGCHRKSVYWEVIGMGTTPEVHSIFLECHTFLVRNH 300
 DB 241 AASARAWPKMTVANGYVNRSLPGLTGCHRKSVYWEVIGMGTTPEVHSIFLECHTFLVRNH 300
 QY 301 RQASLEISPIITFLTAQTLLMDLGOFLFCHSHISSHQHOGNEAYVVKVDSCPHPQRLMKONE 360
 DB 301 RQASLEISPIITFLTAQTLLMDLGOFLFCHSHISSHQHOGNEAYVVKVDSCPHPQRLMKONE 360
 QY 361 EAEDYDDDLTSEMDVVRFDNNSPFIQIRSVAKKPKTWVHYIAAEEEDWDVAPLVLA 420
 DB 361 EAEDYDDDLTSEMDVVRFDNNSPFIQIRSVAKKPKTWVHYIAAEEEDWDVAPLVLA 420
 QY 421 PDRSYKSOYLNNGPQIRGRYKVKVRFMAVYTDFTKTRAIQHESGILGPLLYGSGDTL 480
 DB 421 PDRSYKSOYLNNGPQIRGRYKVKVRFMAVYTDFTKTRAIQHESGILGPLLYGSGDTL 480
 QY 481 LIIFKQASRPYNIYPHGITDVRPLYSRRLPKGVKHLKDFPILPGEIFKYKWTIVVEDGP 540
 DB 481 LIIFKQASRPYNIYPHGITDVRPLYSRRLPKGVKHLKDFPILPGEIFKYKWTIVVEDGP 540
 QY 541 TKSDPRCLTRYSSVNMERDLASGLIGPLLI CYKESVDQGNQIMSKRVILSVDE 600
 DB 541 TKSDPRCLTRYSSVNMERDLASGLIGPLLI CYKESVDQGNQIMSKRVILSVDE 600
 QY 601 NRSWLTENIQRFLEPNAGVQLEDPEFOASNMHSINGYVFDLSQLSVCLHEVAVWYILS 660
 DB 601 NRSWLTENIQRFLEPNAGVQLEDPEFOASNMHSINGYVFDLSQLSVCLHEVAVWYILS 660
 QY 661 IQAQTDFLSVPSGTYFKHKVYEDTLTLPPSGTVMFMSMENGLWILGCHNSDFRNG 720
 DB 661 IQAQTDFLSVPSGTYFKHKVYEDTLTLPPSGTVMFMSMENGLWILGCHNSDFRNG 720
 QY 721 MTALLKVSCKNTQDYYEDYSVEDISAYLLSKNNAIEPRSFQNSRHPSTROKQFNATTI 780
 DB 721 MTALLKVSCKNTQDYYEDYSVEDISAYLLSKNNAIEPRSFQNSRHPSTROKQFNATTI 780
 QY 781 PENDIEKTDVFAHRTMPMKI QNVSSDLLMLLROSPTPHGLSLSDLOBAKYETTFDDPS 840
 DB 781 PENDIEKTDVFAHRTMPMKI QNVSSDLLMLLROSPTPHGLSLSDLOBAKYETTFDDPS 840
 QY 841 PGALDSNNLSLSEMTFRPOLHSGDMVFTPESGLOLRNKLGTAAATLKLDFKVSST 900
 DB 841 PGALDSNNLSLSEMTFRPOLHSGDMVFTPESGLOLRNKLGTAAATLKLDFKVSST 900
 QY 901 SNNLISTIPSDNLAAGTDMTSSLGPPSPVHYDSQDITLTPCKKSPSPLETSGGSLSSE 960
 DB 901 SNNLISTIPSDNLAAGTDMTSSLGPPSPVHYDSQDITLTPCKKSPSPLETSGGSLSSE 960
 QY 961 NNDKLLSGLMNSQBSKGVKSVSTESGRLPKGRAGHPALLTKONALFKYSISLTKN 1020
 DB 961 NNDKLLSGLMNSQBSKGVKSVSTESGRLPKGRAGHPALLTKONALFKYSISLTKN 1020
 QY 1021 KYSNNSATNRKTHIDGPSLLIENSPSVWONILESDFPKKVTPLIHDRMLDKNATALEL 1080
 DB 1021 KYSNNSATNRKTHIDGPSLLIENSPSVWONILESDFPKKVTPLIHDRMLDKNATALEL 1080
 QY 1081 NMSNKTSSKNMVMQCKKEGFI PPDAONPDMSFPFKMLFLPESARWIQRTHGKNSLNSG 1140
 DB 1081 NMSNKTSSKNMVMQCKKEGFI PPDAONPDMSFPFKMLFLPESARWIQRTHGKNSLNSG 1140
 QY 1141 QGSPKQLVSLGPEKVEGQNFLESEKNKVVVGKGEFTKQVGLKEMVFPSSRLFTNLND 1200
 DB 1141 QGSPKQLVSLGPEKVEGQNFLESEKNKVVVGKGEFTKQVGLKEMVFPSSRLFTNLND 1200
 QY 1201 LHENNTNOKKIQBBI EKKETLIQENVVLPOIHTVTGTGKNFMKNLFLLLSTRQNVESYD 1260

Db 1201 L HENNTNORKEKIQEELEK3ETLQENVVLQIHVTGTGKNFMNLFLLSTRQNVESYD 1260
Qy 1261 GAYAPVLQDFPSLNDSTNRTKHTAHFSKKGBEENLEGLGNQTKQIVKCYACTTRISPNT 1320
Db 1261 GAYAPVLQDFPSLNDSTNRTKHTAHFSKKGBEENLEGLGNQTKQIVKCYACTTRISPNT 1320
Qy 1321 SQQNFVQRSKRALQKPLPLEETLEKRIIVDDTSTQWSKNMKHLPSTLTQIDYNEKE 1380
Db 1321 SQQNFVQRSKRALQKPLPLEETLEKRIIVDDTSTQWSKNMKHLPSTLTQIDYNEKE 1380
Qy 1381 KGAITQSPSLCLTRSHSIPQANRSPPIAKVSGPPSIRPIYLTRVLFPQDNSSHLPAASY 1440
Db 1381 KGAITQSPSLCLTRSHSIPQANRSPPIAKVSGPPSIRPIYLTRVLFPQDNSSHLPAASY 1440
Qy 1441 RKDGGVQESSHFQGAQKNNLSAILTLEMTGQREVSGISGTATNSVTYKKVENTVLP 1500
Db 1441 RKDGGVQESSHFQGAQKNNLSAILTLEMTGQREVSGISGTATNSVTYKKVENTVLP 1500
Qy 1501 KPDLPKTSCKVELLPKVHIYOKDLFPPTSTNGSPGHLDVSGSLLOQTGEGAIKNEANRP 1560
Db 1501 KPDLPKTSCKVELLPKVHIYOKDLFPPTSTNGSPGHLDVSGSLLOQTGEGAIKNEANRP 1560
Qy 1561 GKVPFLRVATESSAKTPSKLDPLAWDNHYGTQIPKEEWKQESKPEKTAFFKXDTILSL 1620
Db 1561 GKVPFLRVATESSAKTPSKLDPLAWDNHYGTQIPKEEWKQESKPEKTAFFKXDTILSL 1620
Qy 1621 NACSNHAIAINEQONKPEIEVTWAKQGRTERLCSQNPVULKRHQREITRTTLOSQDEE 1680
Db 1621 NACSNHAIAINEQONKPEIEVTWAKQGRTERLCSQNPVULKRHQREITRTTLOSQDEE 1680
Qy 1681 IDYDITISVEMKEDFDYDEENOSPSFOKTRHYFIAAVERLWDYGMSSPHVLNRR 1740
Db 1681 IDYDITISVEMKEDFDYDEENOSPSFOKTRHYFIAAVERLWDYGMSSPHVLNRR 1740
Qy 1741 AQSGSVPOFKKVVFQFTDGSFTQPLRGELNEHLGLGPIYRAEVEDNIMVTFRNQASR 1800
Db 1741 AQSGSVPOFKKVVFQFTDGSFTQPLRGELNEHLGLGPIYRAEVEDNIMVTFRNQASR 1800
Qy 1801 PYSFYSSLIISYEEDROGAEPKRVKNETKTYFKVQHCHNAPTKDBDFCKAWAYFSDV 1860
Db 1801 PYSFYSSLIISYEEDROGAEPKRVKNETKTYFKVQHCHNAPTKDBDFCKAWAYFSDV 1860
Qy 1861 DLEKDVHSLGLPLVCHTNLPAHGRQVTVQEPALFTTIDETKSWYFTENMERNCRA 1920
Db 1861 DLEKDVHSLGLPLVCHTNLPAHGRQVTVQEPALFTTIDETKSWYFTENMERNCRA 1920
Qy 1921 PCNIQMEDPTFKENYRFAHNGYIMDTLPGLVMAQDQRIRWYLLSGMGNENIHSIHFSGH 1980
Db 1921 PCNIQMEDPTFKENYRFAHNGYIMDTLPGLVMAQDQRIRWYLLSGMGNENIHSIHFSGH 1980
Qy 1981 VFTVKKKEEYKMAJYNYLPQGVFTEVEMLPKAGIWRVECLIGELHLAGNSTLFLVYSNKC 2040
Db 1981 VFTVKKKEEYKMAJYNYLPQGVFTEVEMLPKAGIWRVECLIGELHLAGNSTLFLVYSNKC 2040
Qy 2041 QTPLGWASGHIRDFOITASQYQGWAPKIALRLHYSGSINAWSTKEPFSWKVDLLAPMII 2100
Db 2041 QTPLGWASGHIRDFOITASQYQGWAPKIALRLHYSGSINAWSTKEPFSWKVDLLAPMII 2100
Qy 2101 HGIKTQGARQKFSYISYISOFIIMYSLDGKKWQTYRGNSTGTTLMVFFGNVDSGSIKHINFPN 2160
Db 2101 HGIKTQGARQKFSYISYISOFIIMYSLDGKKWQTYRGNSTGTTLMVFFGNVDSGSIKHINFPN 2160
Qy 2161 PPIIARYIRLPHTHYSIRSTLRMELMGCDLNSCMLPGSKAISDAQTASSYFTNMEFA 2220
Db 2161 PPIIARYIRLPHTHYSIRSTLRMELMGCDLNSCMLPGSKAISDAQTASSYFTNMEFA 2220
Qy 2221 TWSPSKARLHLOGRSNAPQVANNPKZWLOVPQKTMKVGTVTQGVKSLTSMYVKEFL 2280
Db 2221 TWSPSKARLHLOGRSNAPQVANNPKZWLOVPQKTMKVGTVTQGVKSLTSMYVKEFL 2280
Qy 2281 ISSSQDGHQWTLFFQNGKRVKVFQGNQDSFTPVVNSLDPLLTRYLRIHPQSWVHQIALRM 2340

Db 2281 ISSSQDGHQWTLFFQNGKRVKVFQGNQDSFTPVVNSLDPLLTRYLRIHPQSWVHQIALRM 2340
Qy 2341 EVLGCBAQDLY 2351
Db 2341 EVLGCBAQDLY 2351
RESULT 6
ABB99528
ID ABB99528 standard; protein: 2351 AA.
XX ABB99528;
AC ABB99528;
XX 03-MAR-2003 (first entry)
XX Amino acid sequence of human Factor VIII.
DE Factor VIII; non-viral vesicle vector; vesicular membrane; hepatitis B;
XX envelope protein; Factor IX; liver cell; gene therapy; haemophilia.
KW Homo sapiens.
XX WO2002286091-A2.
PN 31-OCT-2002.
XX 25-APR-2002; 2002WO-US013164.
PF 25-APR-2001; 2001US-0286314P.
PR (REGC) UNIV CALIFORNIA.
XX Chien KR, Hoshijima M;
XX WPI; 2003-093125/08.
DR N-PSDB; ABV77052.
XX New non-viral vesicle vector comprises vesicular membrane with hepatitis B envelope protein and nucleic acid expression construct comprising complete factor VIII or IX coding sequence, useful for treating hemophilia.
PS Disclosure; Page 22-32; 34pp; English.
XX The present sequence represents a human Factor VIII. Factor VIII polynucleotides are used to construct non-viral vesicle vectors. These vectors comprise a vesicular membrane with hepatitis B envelope protein exposed on the vesicle surface, and a nucleic acid expression construct comprising a complete factor VIII or factor IX coding sequence and a promoter sequence functional in liver cells. The non-viral vesicle vector is useful in gene therapy protocols for treating haemophilia
XX Sequence 2351 AA;
SQ
Query Match 100.0%; Score 12410; DB 6; Length 2351;
Best Local Similarity 100.0%; Pred.No. 0;
Matches 2350; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 1 MOIELSTCFELCLRFCSATRRYYLGAVELSWDVMQSDLGELPYDARPPRPVKSPFPN 60
Db 1 MOIELSTCFELCLRFCSATRRYYLGAVELSWDVMQSDLGELPYDARPPRPVKSPFPN 60
Qy 61 TSVVYKTLFVEFTDHLFNIAPRPPWMLGPTTQAEVYDTVTILKTMASHPVSLHAV 120
Db 61 TSVVYKTLFVEFTDHLFNIAPRPPWMLGPTTQAEVYDTVTILKTMASHPVSLHAV 120
Qy 121 GVSYWASSEGAFYDQTSORKEKDDKVPFGSGHTTVWQVLKENGPMASDPLCLTYSYLH 180
Db 121 GVSYWASSEGAFYDQTSORKEKDDKVPFGSGHTTVWQVLKENGPMASDPLCLTYSYLH 180
Qy 181 VDLVKDLNSGLIGALLVCREGLAKEKTQTLHKFTLLFAVPDEGKSWHSETKNSLMQDRD 240
Db 181 VDLVKDLNSGLIGALLVCREGLAKEKTQTLHKFTLLFAVPDEGKSWHSETKNSLMQDRD 240

QY 241 AASARAWPMHTVNGYVARSJPLGLGCHRSKYVWHVIGWGTTPVHSHFLECHTFLVRNH 300
DB 241 AASARAWPMHTVNGYVARSJPLGLGCHRSKYVWHVIGWGTTPVHSHFLECHTFLVRNH 300
QY 301 ROASLEISPIITLTATQTLMDLQFLLSCHLSSHQDGMAYVYKVSCEPQLRMKNE 360
DB 301 ROASLEISPIITLTATQTLMDLQFLLSCHLSSHQDGMAYVYKVSCEPQLRMKNE 360
QY 361 EABDYDDDLTDSMDVPRDDDNSSPFTQIRSAVKHFKTWHTWYIAAABEDWDYAPLVLA 420
DB 361 EABDYDDDLTDSMDVPRDDDNSSPFTQIRSAVKHFKTWHTWYIAAABEDWDYAPLVLA 420
QY 421 PDRSVKSOYLANGPQRIQKVKYKVRPMAYTDETFKTREAIQHESGILGLPLYGVBGDTL 480
DB 421 PDRSVKSOYLANGPQRIQKVKYKVRPMAYTDETFKTREAIQHESGILGLPLYGVBGDTL 480
QY 481 LIIFXNQASRPNIYPHGTDVRLYSRRLPKGVKHLKDFPILPGBIFKYKWTVTVEDGP 540
DB 481 LIIFXNQASRPNIYPHGTDVRLYSRRLPKGVKHLKDFPILPGBIFKYKWTVTVEDGP 540
QY 541 TKSDPRCLTRYSSYFNMERDLASGLIGPLLIICYKESVDQGNQIMSDXRNWILSVFDE 600
DB 541 TKSDPRCLTRYSSYFNMERDLASGLIGPLLIICYKESVDQGNQIMSDXRNWILSVFDE 600
QY 601 NRSWYLTENIQRFPLNPAGVQLEDPEFOASNMHSINGVYVDSLOLSVCLHEVAYWYILS 660
DB 601 NRSWYLTENIQRFPLNPAGVQLEDPEFOASNMHSINGVYVDSLOLSVCLHEVAYWYILS 660
QY 661 IGAQTDFLSVFSGYTFKHVMYEDTLTLPFSGETVFMNMNPGWLILGCHNSPFRNG 720
DB 661 IGAQTDFLSVFSGYTFKHVMYEDTLTLPFSGETVFMNMNPGWLILGCHNSPFRNG 720
QY 721 YNALIKVSSCDKNTGNYEDSIEDISAYLLSKNALEPRSPFSONSHPSFTRQKNATTI 780
DB 721 YNALIKVSSCDKNTGNYEDSIEDISAYLLSKNALEPRSPFSONSHPSFTRQKNATTI 780
QY 781 PENDIEKTPWFAHRTPMKIQNVSSDLMILLRQSPFPHGLSLSLQAKYETTSDDPS 840
DB 781 PENDIEKTPWFAHRTPMKIQNVSSDLMILLRQSPFPHGLSLSLQAKYETTSDDPS 840
QY 841 PGAIIDNNLSLSEMTHEPPLHSGDMVFTPEGLQLRLNEKLGTTAATELKKLDFKVSST 900
DB 841 PGAIIDNNLSLSEMTHEPPLHSGDMVFTPEGLQLRLNEKLGTTAATELKKLDFKVSST 900
QY 901 SNKLSTIPSDNLAAGTDNTSSIGPSPMPVHYDSQDITLFGKSSPLTSEGGLSLSEE 960
DB 901 SNKLSTIPSDNLAAGTDNTSSIGPSPMPVHYDSQDITLFGKSSPLTSEGGLSLSEE 960
QY 961 NNDKSLLESGLMNSQSSWGKNVSSSTESGRLFKGKAHGPALLTKDNALFKVYSISLLKTN 1020
DB 961 NNDKSLLESGLMNSQSSWGKNVSSSTESGRLFKGKAHGPALLTKDNALFKVYSISLLKTN 1020
QY 1021 KYSNNSATNRKTHIDGPSLLIENSPEVWQNILESDETFPKVTPFLIHDRMLMDKNATALRL 1080
DB 1021 KYSNNSATNRKTHIDGPSLLIENSPEVWQNILESDETFPKVTPFLIHDRMLMDKNATALRL 1080
QY 1081 NMSNKTTSKJNMWVQKKEGIPDDAQNPDMSFFKMLFLPESARWIORTHGKNSLNG 1140
DB 1081 NMSNKTTSKJNMWVQKKEGIPDDAQNPDMSFFKMLFLPESARWIORTHGKNSLNG 1140
QY 1141 QGSPKQVLSLGPESKVEGQNFISEXNKVVVGKEFTKQVGLKEMVFPSSRNFLFTNLN 1200
DB 1141 QGSPKQVLSLGPESKVEGQNFISEXNKVVVGKEFTKQVGLKEMVFPSSRNFLFTNLN 1200
QY 1201 LHENNTNQBKKIOBIEKKEKTHILOENVLPQHTVTGTFKFNKMLFLLSLSTRQWEGSYD 1260
DB 1201 LHENNTNQBKKIOBIEKKEKTHILOENVLPQHTVTGTFKFNKMLFLLSLSTRQWEGSYD 1260
QY 1261 GAYAPVLQDFRSNDSTNRKKTATHFSSKKEBENLEGLNQTKQIVKQVACTRISPN 1320
DB 1261 GAYAPVLQDFRSNDSTNRKKTATHFSSKKEBENLEGLNQTKQIVKQVACTRISPN 1320

QY 1321 SQQNFVTRSRALKQFRLPLEETELEKRIIVDDTSIQWSKNMKHLTPSTLTQIDYNEKE 1380
DB 1321 SQQNFVTRSRALKQFRLPLEETELEKRIIVDDTSIQWSKNMKHLTPSTLTQIDYNEKE 1380
QY 1381 KGALTQSPLSCLTRSHSIPOANRSPPIAKVSSPSPRIPIYLTRVLPODNSSHLPAASY 1440
DB 1381 KGALTQSPLSCLTRSHSIPOANRSPPIAKVSSPSPRIPIYLTRVLPODNSSHLPAASY 1440
QY 1441 RKQSGVQESSHFIQGAKKNNLSAILTLEMTGDOREVSGLSGTATNSVTYKVKVENTVLP 1500
DB 1441 RKQSGVQESSHFIQGAKKNNLSAILTLEMTGDOREVSGLSGTATNSVTYKVKVENTVLP 1500
QY 1501 KPDLPTKSGKVELLPKVHIYQKDLFPTTSGSPGHLDLVEGSLQGTGAIKWNEANRP 1560
DB 1501 KPDLPTKSGKVELLPKVHIYQKDLFPTTSGSPGHLDLVEGSLQGTGAIKWNEANRP 1560
QY 1561 GKVPFLRVATESSAKTPSKILDPDPLAWDNHYGTQIPKESWKSOEKSPEKTAKKXDTILSL 1620
DB 1561 GKVPFLRVATESSAKTPSKILDPDPLAWDNHYGTQIPKESWKSOEKSPEKTAKKXDTILSL 1620
QY 1621 NACSNHAIAAINEGONKPEIEVTWAKQRTERLCSQNPVVKHQRREITRTTLOSDOEE 1680
DB 1621 NACSNHAIAAINEGONKPEIEVTWAKQRTERLCSQNPVVKHQRREITRTTLOSDOEE 1680
QY 1681 IDYDDTISVEMKEDFDIYDEDNQSPSPFOKKTSHYFIAAVERLWDYGMSSSPHVLNR 1740
DB 1681 IDYDDTISVEMKEDFDIYDEDNQSPSPFOKKTSHYFIAAVERLWDYGMSSSPHVLNR 1740
QY 1741 AQSGSVPOPKVWFOEFTDGSFTQPLVGEHNEHGLGPGYIRAEVEDNIMVTFNQASR 1800
DB 1741 AQSGSVPOPKVWFOEFTDGSFTQPLVGEHNEHGLGPGYIRAEVEDNIMVTFNQASR 1800
QY 1801 PYSYSSLIISYEDDORQGAEPKNEKTYFVKYQHMAPTKDFDCKAWAYFSDV 1860
DB 1801 PYSYSSLIISYEDDORQGAEPKNEKTYFVKYQHMAPTKDFDCKAWAYFSDV 1860
QY 1861 DLEKDVHSLGILGPLLVCHTNTLNPAGHQRVTVQEPALFPTTIDETKSWYFTENMERNCA 1920
DB 1861 DLEKDVHSLGILGPLLVCHTNTLNPAGHQRVTVQEPALFPTTIDETKSWYFTENMERNCA 1920
QY 1921 PCNIQMEDPTKENYRPHAINGYIMDTLPGVMAODQIRWYLLSMGSENIHSHFSGH 1980
DB 1921 PCNIQMEDPTKENYRPHAINGYIMDTLPGVMAODQIRWYLLSMGSENIHSHFSGH 1980
QY 1981 VFTVRKGEYKMALNLYPGVETVEMLPKSAGIWRVECLIGEHLHAGMSTLFLVYSNKC 2040
DB 1981 VFTVRKGEYKMALNLYPGVETVEMLPKSAGIWRVECLIGEHLHAGMSTLFLVYSNKC 2040
QY 2041 QTPLGASGHIRDFOITASGOYGWAPKLARLUHSGSINAWSTKBPFSWIKVDLLAPMII 2100
DB 2041 QTPLGASGHIRDFOITASGOYGWAPKLARLUHSGSINAWSTKBPFSWIKVDLLAPMII 2100
QY 2101 HGIKTQGAQKFSLSYISQFIIMYSIDGKKWOTYRGNSTGTLMVFFGNVDSGIGHNIFN 2160
DB 2101 HGIKTQGAQKFSLSYISQFIIMYSIDGKKWOTYRGNSTGTLMVFFGNVDSGIGHNIFN 2160
QY 2161 PIILARYIRLHPHTHYSIRSLTEMLMGCDLNSCSPLGMESKAISDAQITASSYFTNMPA 2220
DB 2161 PIILARYIRLHPHTHYSIRSLTEMLMGCDLNSCSPLGMESKAISDAQITASSYFTNMPA 2220
QY 2221 TWSPKARLHLOGRNAPVQNNPEKMIQVDFQTKMKTGVTTQGVKSLTSMVVKPEL 2280
DB 2221 TWSPKARLHLOGRNAPVQNNPEKMIQVDFQTKMKTGVTTQGVKSLTSMVVKPEL 2280
QY 2281 ISSODGCHWTLFFONGKVKVFCGNQDSFTPVVNSLDPLLTRYLRIHPQSWHQAIALRM 2340
DB 2281 ISSODGCHWTLFFONGKVKVFCGNQDSFTPVVNSLDPLLTRYLRIHPQSWHQAIALRM 2340
QY 2341 EYLGCEAODLY 2351
DB 2341 EYLGCEAODLY 2351

RESULT 7

AAW11416
 XX AAW11416 standard; protein; 2351 AA.
 AC AAW11416;
 DT 20-NOV-1997 (first entry)
 XX
 DE Active Factor VIII:C analogue A774X.
 KW Factor VIII:C; analogue; glycoprotein; blood coagulation cascade;
 KW fibrinogen; fibrin clot; haemostasis; haemophilia A; bleeding diathesis;
 KW plasma protease; thrombin; immunogen; antibody; haemophiliac; therapy;
 KW proteolytic cleavage.
 XX
 OS Homo sapiens.
 OS Synthetic.
 FH
 FT Key Location/Qualifiers
 FT Peptide 1..19
 FT /note= "signal peptide"
 FT Protein 20..2351
 FT /note= "mature Factor VIII:C"
 FT Region 20..1667
 FT /note= "heavy chain fragment"
 FT Domain 760..1667
 FT /note= "B domain"
 FT Modified-site 774
 FT /label= Phe, Glu, Pro
 FT Region 1668..2350
 FT /note= "light chain fragment"
 XX
 DN W03703195-A1.
 XX
 PD 30-JAN-1997.
 XX
 PF 09-JUL-1996; 96WO-US011444.
 XX
 PR 11-JUL-1995; 95US-0001025P.
 XX
 PA (CHIR) CHIRON CORP.
 XX
 PI Hung DT, Cohen FE, Innis M;
 XX
 XX WPI; 1997-119050/11.
 DR
 XX Factor VIII:C analog modified adjacent to a non-activating Arg residue -
 PT used in the treatment of haemophiliacs, by improvement of haemostasis.
 PT
 XX Claim 26; Page; 90pp; English.
 PS
 CC AAW11330-W11472 represent active Factor VIII:C analogues of the
 CC invention.. These sequences were created by mutating the wild type Factor
 CC VIII:C coding sequence (see AAW1357) using mutagenic primers. The
 CC analogues comprise a native Factor VIII:C polypeptide modified at a site
 CC adjacent to a non-activating Arg residue so that a Arg-Pro or Pro-Arg
 CC dipeptide is created. Factor VIII:C is a large glycoprotein that
 CC participates in the blood coagulation cascade that ultimately converts
 CC soluble fibrinogen to insoluble fibrin clot, effecting haemostasis. A
 CC deficiency in Factor VIII:C is responsible for haemophilia A, which is an
 CC X-chromosome-linked inherited bleeding diathesis. Factor VIII:C is
 CC activated by plasma proteases, such as thrombin. During activation the
 CC mature polypeptide is cleaved to generate heavy and light chain fragments
 CC that are further cleaved. Complexes of two or more of the analogues,
 CC nucleic acids and vectors encoding them may be used alone or in
 CC conjunction with each other, for the prevention or treatment of active
 CC Factor VIII:C deficiency in a mammal. The analogues may be used as
 CC immunogens to raise antibodies, and in the treatment of haemophiliacs, by
 CC improvement of haemostasis. The analogues are resistant to proteolytic
 CC cleavage and display increased plasma half-life. They may be administered
 CC at lower dosages and by different modes of administration
 XX
 SQ Sequence 2351 AA;

Query Match 99.9%; Score 12406; DB 2; Length 2351;
 Best Local Similarity 99.9%; Pred No. 0;
 Matches 2349; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 1 MQIEIETCFCLLRFCFSATRRYYLGAVELSWDMQSDLGELPVDARFPRVPKSPFFN 60
 DB 1 MQIEIETCFCLLRFCFSATRRYYLGAVELSWDMQSDLGELPVDARFPRVPKSPFFN 60
 QY 61 TSVVYKTLFVEFTDHLFNIAKPRPPWMLGPTIOAEVVDVTVITLKNASHVSLHAV 120
 DB 61 TSVVYKTLFVEFTDHLFNIAKPRPPWMLGPTIOAEVVDVTVITLKNASHVSLHAV 120
 QY 121 GVSYWKASEGAEYDDQTSQREKDDKVPFGSSHYYVQVLKENGPMASDPLCLTVSYLSH 180
 DB 121 GVSYWKASEGAEYDDQTSQREKDDKVPFGSSHYYVQVLKENGPMASDPLCLTVSYLSH 180
 QY 181 VDLVKDNLNSGLIGALLVCREGLAKETQTLHKFILLFAVEDEKSKHSETKNSLMQORD 240
 DB 181 VDLVKDNLNSGLIGALLVCREGLAKETQTLHKFILLFAVEDEKSKHSETKNSLMQORD 240
 QY 241 AASARAWPKMTVNGYVNRSLPGLIGCHRKSVYWHVIGMGTTPVHSHFLECHTFLVRNH 300
 DB 241 AASARAWPKMTVNGYVNRSLPGLIGCHRKSVYWHVIGMGTTPVHSHFLECHTFLVRNH 300
 QY 301 ROASLEISPIITFLTAQTLLMDLGOLFLLSCHTSSHOHDGMEAYKVDSCEPEPQOLRMKNE 360
 DB 301 ROASLEISPIITFLTAQTLLMDLGOLFLLSCHTSSHOHDGMEAYKVDSCEPEPQOLRMKNE 360
 QY 361 EAEYDDDLTDSEMDVVRFDNNSPSFTQISVAKKHPKTHWYIAAEEEDWDZAPLVLA 420
 DB 361 EAEYDDDLTDSEMDVVRFDNNSPSFTQISVAKKHPKTHWYIAAEEEDWDZAPLVLA 420
 QY 421 PDDRSYKSOYLNGEPORIGRKVKVRMEAYTDEFKTEATQHSGLIGLLYGVGDTL 480
 DB 421 PDDRSYKSOYLNGEPORIGRKVKVRMEAYTDEFKTEATQHSGLIGLLYGVGDTL 480
 QY 481 LIIFKNQASRPYNIYPHGITDVRPLYSRRLPKGVKHLKDFPILPGEIFKYKWTVTVEDGP 540
 DB 481 LIIFKNQASRPYNIYPHGITDVRPLYSRRLPKGVKHLKDFPILPGEIFKYKWTVTVEDGP 540
 QY 541 TKSDPRCLTRYSSPVNMRDLAAGLIGPLIICYKESVDQGNQIMSKRNVILFSVDE 600
 DB 541 TKSDPRCLTRYSSPVNMRDLAAGLIGPLIICYKESVDQGNQIMSKRNVILFSVDE 600
 QY 601 NESWYLTENIQRFLENPAGVQLEDPEFOASNMHSINGVVFDSQLSVCLHEVAYWYILS 660
 DB 601 NESWYLTENIQRFLENPAGVQLEDPEFOASNMHSINGVVFDSQLSVCLHEVAYWYILS 660
 QY 661 IGAQTDFLSVFPGSYTFKHQWYEDTLTLPFPGSETVFMSEMPGLWILGCHNSDFNRG 720
 DB 661 IGAQTDFLSVFPGSYTFKHQWYEDTLTLPFPGSETVFMSEMPGLWILGCHNSDFNRG 720
 QY 721 MTALLKVSCKNTGDYEDSDYEDISAYLLSKNNAIEPRSPQNSRHSHTROKQNAFTI 780
 DB 721 MTALLKVSCKNTGDYEDSDYEDISAYLLSKNNAIEPRSPQNSRHSHTROKQNAFTI 780
 QY 781 PENDIEKTDPMFAHETPEKIQNVSSDILLMLROSPTPHGLSLSDLOEAKYETESDDPS 840
 DB 781 PENDIEKTDPMFAHETPEKIQNVSSDILLMLROSPTPHGLSLSDLOEAKYETESDDPS 840
 QY 841 PGADSNNSLSEMTFRPQLHSGDMVTTPRSGLQRLNEKLGTTAATLKKLDFKVSST 900
 DB 841 PGADSNNSLSEMTFRPQLHSGDMVTTPRSGLQRLNEKLGTTAATLKKLDFKVSST 900
 QY 901 SNNLSTIPSDNLAAGTNTSSLGPPSPVHYDSOLDITLPGKKSPLTESGPPSLSEE 960
 DB 901 SNNLSTIPSDNLAAGTNTSSLGPPSPVHYDSOLDITLPGKKSPLTESGPPSLSEE 960
 QY 961 NNDSKLLSGLMNSOBSSWGKNVSTESORLFGKARAHGPALLTKDNALFKVSIULLKTN 1020
 DB 961 NNDSKLLSGLMNSOBSSWGKNVSTESORLFGKARAHGPALLTKDNALFKVSIULLKTN 1020

QY 1021 KTSNNSATNRKTHIDGPSILLIENSFSWQNILESDETEKKTPTPLHDMRMDKXATLRLL 1080
DB 1021 KTSNNSAINRKTHIDGPSILLIENSFSWQNILESDETEKKTPTPLHDMRMDKXATLRLL 1080
QY 1081 NEMSNKTTSSKNMVMQOKKEGPIPPDAQNPDMSFFKMLFJPESARWIQTHGKXSLNSG 1140
DB 1081 NEMSNKTTSSKNMVMQOKKEGPIPPDAQNPDMSFFKMLFJPESARWIQTHGKXSLNSG 1140
QY 1141 QGSPKQVLVSLGPESKVEQNFSLSKNVKVGKGFVKDVGKEMVFPSSRNLFITNLN 1200
DB 1141 QGSPKQVLVSLGPESKVEQNFSLSKNVKVGKGFVKDVGKEMVFPSSRNLFITNLN 1200
QY 1201 LHENNTHQEKKIQSEIEBKETLIIQXNVLPQIHVTGFKNFMKMLFLLSTRQNVGSSVD 1260
DB 1201 LHENNTHQEKKIQSEIEBKETLIIQXNVLPQIHVTGFKNFMKMLFLLSTRQNVGSSVD 1260
QY 1261 GAYAPVLQDFRSLNDSTNRKTHAHFSKKGSEENLEGLGNOTKOIYEKIACITRISNPT 1320
DB 1261 GAYAPVLQDFRSLNDSTNRKTHAHFSKKGSEENLEGLGNOTKOIYEKIACITRISNPT 1320
QY 1321 SOQNFVQSSKALKOPRPLPBEETELKELIYDDTSTQWSKNMKHLTPSTLTQIDYNEKE 1380
DB 1321 SOQNFVQSSKALKOPRPLPBEETELKELIYDDTSTQWSKNMKHLTPSTLTQIDYNEKE 1380
QY 1381 KGALTQSPUSDCLTRSHSIPQANRSPPIAKVSSPSPRIPIYLVTRVLFQDSSSHLPAASY 1440
DB 1381 KGALTQSPUSDCLTRSHSIPQANRSPPIAKVSSPSPRIPIYLVTRVLFQDSSSHLPAASY 1440
QY 1441 RKDQSGVQBSHFLOKAKKNNLSLAILTLEMTGDQREVGSLGTSATNSVTYKKVENTVLP 1500
DB 1441 RKDQSGVQBSHFLOKAKKNNLSLAILTLEMTGDQREVGSLGTSATNSVTYKKVENTVLP 1500
QY 1501 KEDLPKTSKVELLPKVHVIYQKDLFPTETSNQSPGHLQVEGSLQGTGCAIKWNEANRP 1560
DB 1501 KEDLPKTSKVELLPKVHVIYQKDLFPTETSNQSPGHLQVEGSLQGTGCAIKWNEANRP 1560
QY 1561 GKVPFLRVATSSAKTPSKLLDPLANDNHYGTQIPKEWKSQEKSPKTAFAKKOTILSL 1620
DB 1561 GKVPFLRVATSSAKTPSKLLDPLANDNHYGTQIPKEWKSQEKSPKTAFAKKOTILSL 1620
QY 1621 NACESNHAIAANEGQNKKEIEVTWAKQGTERTCLSQNPVLRKHOREITRITLQSDQEE 1680
DB 1621 NACESNHAIAANEGQNKKEIEVTWAKQGTERTCLSQNPVLRKHOREITRITLQSDQEE 1680
QY 1681 IDYDDTISVEMKKEDDIYDDEDNQSPGPRFOKTRHYFIAAVERLWDYGNSSSPHVLNR 1740
DB 1681 IDYDDTISVEMKKEDDIYDDEDNQSPGPRFOKTRHYFIAAVERLWDYGNSSSPHVLNR 1740
QY 1741 AQSGSVPPQKVVQFQFTDGSFTQPLRYGSELNEHLGLGPIRAEVEDNIMVTRKQASR 1800
DB 1741 AQSGSVPPQKVVQFQFTDGSFTQPLRYGSELNEHLGLGPIRAEVEDNIMVTRKQASR 1800
QY 1801 PYSFYSLSISYEDORQGAZPRKNFYKPNETKTYFWKVQHMAPTKDEPCKAWAYFSDV 1860
DB 1801 PYSFYSLSISYEDORQGAZPRKNFYKPNETKTYFWKVQHMAPTKDEPCKAWAYFSDV 1860
QY 1861 DLBKDVHSGLIGLLVCHTNTLPAHGRQVTVQEPFALFFTFDETKSWYFTENMERNCR 1920
DB 1861 DLBKDVHSGLIGLLVCHTNTLPAHGRQVTVQEPFALFFTFDETKSWYFTENMERNCR 1920
QY 1921 PCNIQMEDPTFKENYRPHANINGITMDYLPGLVNAQDQRIWYLLSMGSENHISIHFSGH 1980
DB 1921 PCNIQMEDPTFKENYRPHANINGITMDYLPGLVNAQDQRIWYLLSMGSENHISIHFSGH 1980
QY 1981 VFTVRKEEYKALNYLYPGVEFTEVEMLPKAGIWRVECLIGHLAGMSTFLVYSNKC 2040
DB 1981 VFTVRKEEYKALNYLYPGVEFTEVEMLPKAGIWRVECLIGHLAGMSTFLVYSNKC 2040
QY 2041 QTPLGASGHIRDFOITASGOYGOWAPKLAPLHYVSGSINAWSTKEPFSWIKVDLLAPMII 2100
DB 2041 QTPLGASGHIRDFOITASGOYGOWAPKLAPLHYVSGSINAWSTKEPFSWIKVDLLAPMII 2100
QY 2101 HGIKTQARQKSFSSLYISQFIIMTSLDCKKQWYRGNSGTGLMVFVFGNVDSSGIGKINFN 2160

DB 2101 HGIKTQARQKSFSSLYISQFIIMTSLDCKKQWYRGNSGTGLMVFVFGNVDSSGIGKINFN 2160
QY 2161 PPIIARYIRLPHTHYSIRSTLRMBELMGDLNCSMPGLGWSKASDAQITASSYFTNMEA 2220
DB 2161 PPIIARYIRLPHTHYSIRSTLRMBELMGDLNCSMPGLGWSKASDAQITASSYFTNMEA 2220
QY 2221 TWSPSKARLHQGSNARWPOVNNPKEWLQVDFOKTMKVTGVTQGVKSLLTSMYVKEFL 2280
DB 2221 TWSPSKARLHQGSNARWPOVNNPKEWLQVDFOKTMKVTGVTQGVKSLLTSMYVKEFL 2280
QY 2281 ISSQDGHQWTLFFONGKVKVFOQNQGSFTPVVNSLDPPLLTRYLRIHPQSWHQAIALRM 2340
DB 2281 ISSQDGHQWTLFFONGKVKVFOQNQGSFTPVVNSLDPPLLTRYLRIHPQSWHQAIALRM 2340
QY 2341 EVLGEAODLY 2351
DB 2341 EVLGEAODLY 2351
RESULT 8
AAW11343
ID AAW11343 standard; protein; 2351 AA.
XX AC AAW11343;
XX DT 17-NOV-1997 (first entry)
XX DE Active Factor VIII:C analogue S224X.
XX DE Factor VIII:C; analogue; glycoprotein; blood coagulation cascade;
KW fibrinogen; fibrin clot; haemostasis; haemophilia A; bleeding diathesis;
KW plasma protease; thrombin; immunogen; antibody; haemophilic; therapy;
XX proteolytic cleavage.
XX OS Homo sapiens.
XX OS Synthetic.
FH Key
FH Peptide 1. .19 Location/Qualifiers
FT /note= "signal peptide"
FT Protein 20..2351 /note= "mature Factor VIII:C"
FT Region 20..1667 /note= "heavy chain fragment"
FT Modified-site 243 /label= Phe, Glu, Pro
FT Domain 760..1667 /note= "B domain"
FT Region 1668..2350 /note= "light chain fragment"
XX WC9703195-AI.
XX 30-JAN-1997.
XX 09-JUL-1996; 96WO-US011444.
XX 11-JUL-1995; 95US-0001025P.
XX (CHIR) CHIRON CORP.
XX Hung DT, Cohen FE, Innis M;
XX WPI; 1997-119050/11.
XX Factor VIII:C analog modified adjacent to a non-activating Arg residue -
PT used in the treatment of haemophilias; by improvement of haemostasis.
XX Claim 10; Page; 90pp; English.
XX AAW11330-W11472 represent active Factor VIII:C analogues of the
CC invention. These sequences were created by mutating the wild type Factor

CC VIII:C coding sequence (see BAT51357) using mutagenic primers. The
 CC analogues comprise a native Factor VIII:C polypeptide modified at a site
 CC adjacent to a non-activating Arg residue so that a Arg-Pro or Pro-Arg
 CC dipeptide is created. Factor VIII:C is a large glycoprotein that
 CC participates in the blood coagulation cascade that ultimately converts
 CC soluble fibrinogen to insoluble fibrin clot, effecting haemostasis. A
 CC deficiency in Factor VIII:C is responsible for haemophilia A, which is an
 CC X-chromosome-linked inherited bleeding diathesis. Factor VIII:C is
 CC activated by plasma proteases, such as thrombin. During activation the
 CC mature polypeptide is cleaved to generate heavy and light chain fragments
 CC that are further cleaved. Complexes of two or more of the analogues,
 CC nucleic acids and vectors encoding them may be used alone or in
 CC conjunction with each other, for the prevention or treatment of active
 CC Factor VIII:C deficiency in a mammal. The analogues may be used as
 CC immunogens to raise antibodies, and in the treatment of haemophilias, by
 CC improvement of haemostasis. The analogues are resistant to proteolytic
 CC cleavage and display increased plasma half life. They may be administered
 CC at lower dosages and by different modes of administration
 XX
 SQ Sequence 2351 AA;

Query Match 99.9%; Score 2406; DB 2; Length 2351;
 Best Local Similarity 99.9%; Pred. No. 0;
 Matches 2349; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy	1	MOELSTCFCLLAFCEFSATRRYYLGHVLSWDMQSDLGELPVDARFPRVPKSRPPN	60
Db	1	MOELSTCFCLLAFCEFSATRRYYLGHVLSWDMQSDLGELPVDARFPRVPKSRPPN	60
Qy	61	TSVYVYKTLFVEFTDHLFNIAKPRPPWMLGPGTQAQVYTVITLKNMASHPVSLHAV	120
Db	61	TSVYVYKTLFVEFTDHLFNIAKPRPPWMLGPGTQAQVYTVITLKNMASHPVSLHAV	120
Qy	121	GVSWKASEGAEYDQTSQRKEDDKVPPGSGHTYVWQVLKENGPMASDPLCLTVSYLSH	180
Db	121	GVSWKASEGAEYDQTSQRKEDDKVPPGSGHTYVWQVLKENGPMASDPLCLTVSYLSH	180
Qy	181	VDLVKNSGLTGALLVCRSGSLAKEKQTQLHKFILLFAVDEGKSWHSETKNSLMQDRD	240
Db	181	VDLVKNSGLTGALLVCRSGSLAKEKQTQLHKFILLFAVDEGKSWHSETKNSLMQDRD	240
Qy	241	AASAAFAFMHTVNGVYVNRSLGLICGHRKSYVHVIWGMITPEVHSIFLEGHTFLVRNH	300
Db	241	AAXAAFAFMHTVNGVYVNRSLGLICGHRKSYVHVIWGMITPEVHSIFLEGHTFLVRNH	300
Qy	301	QOASLEISPIITFLTAQTLMDLIGQPLLSCHTSSHOHDGMEAYVVDSCPEBPOLRMKNE	360
Db	301	QOASLEISPIITFLTAQTLMDLIGQPLLSCHTSSHOHDGMEAYVVDSCPEBPOLRMKNE	360
Qy	361	EAEDYDDDLTDSMDVRFDDDNSPSFTQIRSAVKHKPKTWHTYTAAREEDWDYAPLVIA	420
Db	361	EAEDYDDDLTDSMDVRFDDDNSPSFTQIRSAVKHKPKTWHTYTAAREEDWDYAPLVIA	420
Qy	421	PDDRYSKYQLNNGPQRIQKRYKVFWAYTDETEKTREAIQHESGILGPLYGEGDTL	480
Db	421	PDDRYSKYQLNNGPQRIQKRYKVFWAYTDETEKTREAIQHESGILGPLYGEGDTL	480
Qy	481	LIIFKQASRPNIYVPHGITDVRPLYSRRLPKGVHLKDPFPLPGELPKYKWTVTVEDGP	540
Db	481	LIIFKQASRPNIYVPHGITDVRPLYSRRLPKGVHLKDPFPLPGELPKYKWTVTVEDGP	540
Qy	541	TKSDPRCLTRYYSFVNMERDLASGLIGPLLTCYKESVDORGNQIMSKRNILSVFDE	600
Db	541	TKSDPRCLTRYYSFVNMERDLASGLIGPLLTCYKESVDORGNQIMSKRNILSVFDE	600
Qy	601	NRSWYLTENIQRFELPNAGVQLEDPEFOASNMHSINGVYVPSLQLSVCLHEVAYWYILS	660
Db	601	NRSWYLTENIQRFELPNAGVQLEDPEFOASNMHSINGVYVPSLQLSVCLHEVAYWYILS	660
Qy	661	ICAQTDPLSVFSGVTFKHKMYEDTLTLFPFSGHTVFMNMPNGLWILGCHNSDFNRNG	720
Db	661	ICAQTDPLSVFSGVTFKHKMYEDTLTLFPFSGHTVFMNMPNGLWILGCHNSDFNRNG	720

Qy	721	MTALLKVSSCKNTGDIYEDSYEDI SAYLLSKNNAIEPRSFQNSRHPSTROKQFNATTI	780
Db	721	MTALLKVSSCKNTGDIYEDSYEDI SAYLLSKNNAIEPRSFQNSRHPSTROKQFNATTI	780
Qy	781	PENDIEXTDWFPAHRTMPMKIQNVSSDILLMLLRQSPTRHGLSLSDIQEAKYETFSDDPS	840
Db	781	PENDIEXTDWFPAHRTMPMKIQNVSSDILLMLLRQSPTRHGLSLSDIQEAKYETFSDDPS	840
Qy	841	PGALDSNNSISEMTHFPOLHSGDMVFTBESGLQELNEKLGTTAATELKKLDFKVSST	900
Db	841	PGALDSNNSISEMTHFPOLHSGDMVFTBESGLQELNEKLGTTAATELKKLDFKVSST	900
Qy	901	SNNLITIPSDNLAAAGTNTSSLGPPSPMPVHYDSQDLTTLFGKSSPLTSGGSLISEE	960
Db	901	SNNLITIPSDNLAAAGTNTSSLGPPSPMPVHYDSQDLTTLFGKSSPLTSGGSLISEE	960
Qy	961	NNDKLLSGLMNSQESSWGKNVSTESGRLFKGRAGPALLTKDNALFKVSI SLKTN	1020
Db	961	NNDKLLSGLMNSQESSWGKNVSTESGRLFKGRAGPALLTKDNALFKVSI SLKTN	1020
Qy	1021	KTSNNSATNRKTHIDGPSLLIENSPSVWOKILBSTDTEFKVTPLIHDRMLMDKNATLRL	1080
Db	1021	KTSNNSATNRKTHIDGPSLLIENSPSVWOKILBSTDTEFKVTPLIHDRMLMDKNATLRL	1080
Qy	1081	NHNSKNTTSSKQMEMVQKKEGPPIPPDAQNPDMSFFKMLFLPESARWIQRTHGKNSLNSG	1140
Db	1081	NHNSKNTTSSKQMEMVQKKEGPPIPPDAQNPDMSFFKMLFLPESARWIQRTHGKNSLNSG	1140
Qy	1141	QGPSFKQIVSLGPKSVGEGQNFISEKNVVGKGEFTQVGLKEMVFPSSNLFITNLDN	1200
Db	1141	QGPSFKQIVSLGPKSVGEGQNFISEKNVVGKGEFTQVGLKEMVFPSSNLFITNLDN	1200
Qy	1201	LHENNTNHOEKKIOEIEKKETLLOENNVLPQIHTVTCTKXFMKNLFLSTROKVEGSD	1260
Db	1201	LHENNTNHOEKKIOEIEKKETLLOENNVLPQIHTVTCTKXFMKNLFLSTROKVEGSD	1260
Qy	1261	GAYAPVLQDFPSLNDSTNRKTHAHFSKKEENLEGNGQTQKQIVKAYACTRISNT	1320
Db	1261	GAYAPVLQDFPSLNDSTNRKTHAHFSKKEENLEGNGQTQKQIVKAYACTRISNT	1320
Qy	1321	SQQNFTQRSKRALKQFRLPLEETELEKRIIVDDTSTOWSKNMKHLTPSTLTQIDYNEKE	1380
Db	1321	SQQNFTQRSKRALKQFRLPLEETELEKRIIVDDTSTOWSKNMKHLTPSTLTQIDYNEKE	1380
Qy	1381	KGATQSPSLDCLFRSHSIPOANESPLIAKVSPSPSRPIYLTREVLFDQKSSHLPAASY	1440
Db	1381	KGATQSPSLDCLFRSHSIPOANESPLIAKVSPSPSRPIYLTREVLFDQKSSHLPAASY	1440
Qy	1441	RKDSGVQESSHFTQGAQKNNLSLAITLLEMTGQREVSGSLGTSATNSVTYKVENTVLP	1500
Db	1441	RKDSGVQESSHFTQGAQKNNLSLAITLLEMTGQREVSGSLGTSATNSVTYKVENTVLP	1500
Qy	1501	KPDLPKTSKVELLPKHIIYOKDLFPPTSETNGSGHLDLVEGSLLOGTEGAIKNEANRP	1560
Db	1501	KPDLPKTSKVELLPKHIIYOKDLFPPTSETNGSGHLDLVEGSLLOGTEGAIKNEANRP	1560
Qy	1561	GKVPFLRVATESSAKTPSKLLDPLAWDNHYGTQIPKEEWKQEKSPKETAFAKKDITLSL	1620
Db	1561	GKVPFLRVATESSAKTPSKLLDPLAWDNHYGTQIPKEEWKQEKSPKETAFAKKDITLSL	1620
Qy	1621	NACSNHAI AAINSGQNKPEIEVTWAKGRTERLCSQNPVVLKRHOREITRTTLQSDQEE	1680
Db	1621	NACSNHAI AAINSGQNKPEIEVTWAKGRTERLCSQNPVVLKRHOREITRTTLQSDQEE	1680
Qy	1681	IDYDDTISVENKKGDFFIYDEDENQSPRSFOKTRHYPIAAVERLWDYGMSSSHVLENR	1740
Db	1681	IDYDDTISVENKKGDFFIYDEDENQSPRSFOKTRHYPIAAVERLWDYGMSSSHVLENR	1740
Qy	1741	AQSSVPOFKVVFQEPDGSFTQPLVYRGEINELHGLLGPVIRAEVDENIMVTFRNQASR	1800
Db	1741	AQSSVPOFKVVFQEPDGSFTQPLVYRGEINELHGLLGPVIRAEVDENIMVTFRNQASR	1800
Qy	1801	PYSYSSLSIYEEDQORQGAEPKQNFVKPNETKTIVFWKYQHNAVPTKEFFDCKAWAYFSDV	1860

```
Db 1801 PYSFYSSLIISXEDRQRAEPRKPVKNEYTKYQVHEHAPKDFDCKAWAYFSDV 1860
QY 1861 DLEKDVHSLGLPLLVCHTNTLNPAHQVQTVQEPALFFTFIDETKSWYFTENMERNCR 1920
Db 1861 DLEKDVHSLGLPLLVCHTNTLNPAHQVQTVQEPALFFTFIDETKSWYFTENMERNCR 1920
QY 1921 PCNIQMEDPTKENYRFAINGYINDTLPLGVMADQRIWVLLSMGSGNENIHSIHPSGH 1980
Db 1921 PCNIQMEDPTKENYRFAINGYINDTLPLGVMADQRIWVLLSMGSGNENIHSIHPSGH 1980
QY 1981 VFTVKKKEEYKMALNLYPGVFTVEMLPFSKAGIWRVCECLIGEHLHAGMSILFVYSNKC 2040
Db 1981 VFTVKKKEEYKMALNLYPGVFTVEMLPFSKAGIWRVCECLIGEHLHAGMSILFVYSNKC 2040
QY 2041 QTPLQMASCHIRDFQITASQYQWAPKLIARLHYSGSINAWSTKSPFSWIKVDLLAPMII 2100
Db 2041 QTPLQMASCHIRDFQITASQYQWAPKLIARLHYSGSINAWSTKSPFSWIKVDLLAPMII 2100
QY 2101 HGIKTQGAQKSFSSLYISQFIIMYSIDGKKWOTYGNSTGTLMVFPFGNVDSGIGKHNFN 2160
Db 2101 HGIKTQGAQKSFSSLYISQFIIMYSIDGKKWOTYGNSTGTLMVFPFGNVDSGIGKHNFN 2160
QY 2161 PPIIARYIRLHPHTHSIRSLRMLMGCDLNSCSPLGMESKAI SDAQITASSYFTNMFA 2220
Db 2161 PPIIARYIRLHPHTHSIRSLRMLMGCDLNSCSPLGMESKAI SDAQITASSYFTNMFA 2220
QY 2221 TWSPSKARLHLOGRGNARFPQVNNPKWLVQDFQKTMKVYGTGTVGVKSLTSMYVKZPL 2280
Db 2221 TWSPSKARLHLOGRGNARFPQVNNPKWLVQDFQKTMKVYGTGTVGVKSLTSMYVKZPL 2280
QY 2281 ISSSQDGHQWTLFFQNGKVKVQGNQDSFTPVVNSLDPPLLTRYLRHPQSWVHQIALRM 2340
Db 2281 ISSSQDGHQWTLFFQNGKVKVQGNQDSFTPVVNSLDPPLLTRYLRHPQSWVHQIALRM 2340
QY 2341 EVLGCEAQDLY 2351
Db 2341 EVLGCEAQDLY 2351

RESULT 9
AAW11435
ID AAW11435 standard; protein; 2351 AA.
AC AAW11435;
XX
DF 20-NOV-1997 (first entry)
XX
DE Active Factor VIII:C analogue S1311X.
XX
KW Factor VIII:C; analogue; glycoprotein; blood coagulation cascade;
KW fibrinogen; fibrin clot; haemostasis; haemophilia A; bleeding diathesis;
KW plasma protease; thrombin; immunogen; antibody; haemophilic; therapy;
KW proteolytic cleavage.
XX
OS Homo sapiens.
OS Synthetic.
XX
FH Key Location/Qualifiers
FT Peptide 1..19
FT Protein /note= "signal peptide"
FT Region /note= "mature Factor VIII:C"
FT Domain /note= "heavy chain fragment"
FT Modified-site 760..1667
FT Region /note= "B domain"
FT Region 1330
FT Region /label= Phe, Glu, Pro
FT Region 1668..2350
FT Region /note= "light chain fragment"
XX
FN MO9703195-A1.
```

```
XX 30-JAN-1997.
PD
XX
PF 09-JUL-1996; 96WO-US011444.
XX
PR 11-JUL-1995; 95US-0001025P.
XX (CHIR ) CHIRON CORP.
XX
PI Hung DT, Cohen FE, Innis M;
XX WPI; 1997-119050/11.
DR
XX
PT Factor VIII:C analog modified adjacent to a non-activating Arg residue -
PT used in the treatment of haemophilias, by improvement of haemostasis.
XX
PS Claim 30; Page; 90pp; English.
XX
CC AAW11330-W11472 represent active Factor VIII:C analogues of the
CC invention. These sequences were created by mutating the wild type Factor
CC VIII:C coding sequence (see AWT51357) using mutagenic primers. The
CC analogues comprise a native Factor VIII:C polypeptide modified at a site
CC adjacent to a non-activating Arg residue so that a Arg-Pro or Pro-Arg
CC dipeptide is created. Factor VIII:C is a large glycoprotein that
CC participates in the blood coagulation cascade that ultimately converts
CC soluble fibrinogen to insoluble fibrin clot, effecting haemostasis. A
CC deficiency in Factor VIII:C is responsible for haemophilia A, which is an
CC X-chromosome-linked inherited bleeding diathesis. Factor VIII:C is
CC activated by plasma proteases, such as thrombin. During activation the
CC mature polypeptide is cleaved to generate heavy and light chain fragments
CC that are further cleaved. Complexes of two or more of the analogues,
CC nucleic acids and vectors encoding them may be used alone or in
CC conjunction with each other, for the prevention or treatment of active
CC Factor VIII:C deficiency in a mammal. The analogues may be used as
CC immunogens to raise antibodies, and in the treatment of haemophilias, by
CC improvement of haemostasis. The analogues are resistant to proteolytic
CC cleavage and display increased plasma half-life. They may be administered
CC at lower dosages and by different modes of administration
XX
SQ Sequence 2351 AA;
Query Match 99.9%; Score 12406; DB 2; Length 2351;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 2349; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 MOELSTCFELCLIRFCFSATRRYYLGAVELSWDYNQSDLGELPVDARPPRPVPSFPFN 60
Db 1 MOELSTCFELCLIRFCFSATRRYYLGAVELSWDYNQSDLGELPVDARPPRPVPSFPFN 60
QY 61 TSVVYKTLFVEFTDHLNIAKPRPPNMGLLGPTIOAEVYDVTWITLKNMASHPVSLHAV 120
Db 61 TSVVYKTLFVEFTDHLNIAKPRPPNMGLLGPTIOAEVYDVTWITLKNMASHPVSLHAV 120
QY 121 GVSVMKASEGAEVDDQTSQREKDDKVPFGSGHTVWQVLKENGPMASDPLCLTYSYLSH 180
Db 121 GVSVMKASEGAEVDDQTSQREKDDKVPFGSGHTVWQVLKENGPMASDPLCLTYSYLSH 180
QY 181 VDLVKDLNSGLIGALLVCREGSLAKEKTQTLHKFILLFAVDFEGKSWHSETKNSLMQDRD 240
Db 181 VDLVKDLNSGLIGALLVCREGSLAKEKTQTLHKFILLFAVDFEGKSWHSETKNSLMQDRD 240
QY 241 AASARAWPKCHTNGVYVNRSLPGLIGCHRKSVYVHVIQMGTTPEVHSIFLEGHTFLVRNH 300
Db 241 AASARAWPKCHTNGVYVNRSLPGLIGCHRKSVYVHVIQMGTTPEVHSIFLEGHTFLVRNH 300
QY 301 ROASLEISPIITELTAQTLLMDLQGLLSCHISHOHGDMGAEVYKVDSCPEEPQLRMKNE 360
Db 301 ROASLEISPIITELTAQTLLMDLQGLLSCHISHOHGDMGAEVYKVDSCPEEPQLRMKNE 360
QY 361 EAEDYDDDLTDGEMDVRFDDDDNSPSFQIRSVAKKHPTWVHYIAABEEDWDYAPLVLA 420
Db 361 EAEDYDDDLTDGEMDVRFDDDDNSPSFQIRSVAKKHPTWVHYIAABEEDWDYAPLVLA 420
```

QY 421 PDRS YKSOYLNNQPORIGSKYKKVREMYATDETEKTRERAIQHSGLIGPLLGEVGDTL 480
DB 421 PDRS YKSOYLNNQPORIGSKYKKVREMYATDETEKTRERAIQHSGLIGPLLGEVGDTL 480
QY 481 LIIFKQASRYNIYPHGIDTVRPLYRRRLPGVKHLKDRPILPFGIEPKYKWTVTVEDGP 540
DB 481 LIIFKQASRYNIYPHGIDTVRPLYRRRLPGVKHLKDRPILPFGIEPKYKWTVTVEDGP 540
QY 541 TKSDPRCLTRYYSFVNMERDLASGLIGPLLICYKESVDORGNOIMSDKRNVLFSVFDE 600
DB 541 TKSDPRCLTRYYSFVNMERDLASGLIGPLLICYKESVDORGNOIMSDKRNVLFSVFDE 600
QY 601 NRSWLTENIQRLPUPNAGVLEDBEFOASNIMESINGYVDFSQLSVCLHEVAYWILS 660
DB 601 NRSWLTENIQRLPUPNAGVLEDBEFOASNIMESINGYVDFSQLSVCLHEVAYWILS 660
QY 661 IGAOTDLSVFFSGYTPFKHKVYEDTLTLPFSGETVFMSENFGMLWILGCHNSDFNRG 720
DB 661 IGAOTDLSVFFSGYTPFKHKVYEDTLTLPFSGETVFMSENFGMLWILGCHNSDFNRG 720
QY 721 MTALLKVSSCDKNTGDIYEDSYEDISAYLLSKNAIAPRSFSONSRHPSRQKQFNATTI 780
DB 721 MTALLKVSSCDKNTGDIYEDSYEDISAYLLSKNAIAPRSFSONSRHPSRQKQFNATTI 780
QY 781 PENDIEKTDPAFAHRTMPKIQNVSSDMLLRQSPHPGLSLSDLOEAKYETFSDDPS 840
DB 781 PENDIEKTDPAFAHRTMPKIQNVSSDMLLRQSPHPGLSLSDLOEAKYETFSDDPS 840
QY 841 PGATDSNNSLSEMTFRPQLHSGDMVFTPSGGIQLRLNEKLGTTAATLKKLDFKYSST 900
DB 841 PGATDSNNSLSEMTFRPQLHSGDMVFTPSGGIQLRLNEKLGTTAATLKKLDFKYSST 900
QY 901 SNNLISITPSNLAAGTNTSGLIPPSKPVHYDSOLDTTLFGKXSSPLTESGGPLSSEE 960
DB 901 SNNLISITPSNLAAGTNTSGLIPPSKPVHYDSOLDTTLFGKXSSPLTESGGPLSSEE 960
QY 961 NNDGKLLBSGLMNSQSSWGKNVSTEGRLFKKRAHGPALLTKDNALFKVSISSLKTN 1020
DB 961 NNDGKLLBSGLMNSQSSWGKNVSTEGRLFKKRAHGPALLTKDNALFKVSISSLKTN 1020
QY 1021 KTSNNSATNRKTHIDCPSELLIENSFVWQNILESDETEFKVTPLIHDRMLMDKXATLRL 1080
DB 1021 KTSNNSATNRKTHIDCPSELLIENSFVWQNILESDETEFKVTPLIHDRMLMDKXATLRL 1080
QY 1081 NMSNKTTSKXMEVQOKKEGPIPDQAQNPMSFFKMLFIPESARWQRTHGKNSLNSG 1140
DB 1081 NMSNKTTSKXMEVQOKKEGPIPDQAQNPMSFFKMLFIPESARWQRTHGKNSLNSG 1140
QY 1141 QCPSPKQVSLGPEKSVEGQNFLEKKNVVGKGFTKDVGLKEMVFPSSRNLFILTNLDN 1200
DB 1141 QCPSPKQVSLGPEKSVEGQNFLEKKNVVGKGFTKDVGLKEMVFPSSRNLFILTNLDN 1200
QY 1201 LHENNTNHOEKIQBIEIKKETLIIQENVVLPQIHVVTGTKNPKNLFILSTRQNVGSGYD 1260
DB 1201 LHENNTNHOEKIQBIEIKKETLIIQENVVLPQIHVVTGTKNPKNLFILSTRQNVGSGYD 1260
QY 1261 GAYAPVLQDFPSLNDSTNRKKGTAHFSKKGSEENLEGLGNQTKQIYKACTTIRISNPT 1320
DB 1261 GAYAPVLQDFPSLNDSTNRKKGTAHFSKKGSEENLEGLGNQTKQIYKACTTIRISNPT 1320
QY 1321 SQONFVQTSKRALQFRLPLEETELEKRIIVDDTSTOWSKNMKHLTPSTLTQIDYNEKE 1380
DB 1321 SQONFVQTSKRALQFRLPLEETELEKRIIVDDTSTOWSKNMKHLTPSTLTQIDYNEKE 1380
QY 1381 KGAIQTOSPLSDCLTRSHSIPQANRSLPIAKVSSFPSPRIPIYLTTRVLFPQDNESHLPAAZY 1440
DB 1381 KGAIQTOSPLSDCLTRSHSIPQANRSLPIAKVSSFPSPRIPIYLTTRVLFPQDNESHLPAAZY 1440
QY 1441 RKXDSGVQSSHFLOGAKKNLSLALTLEMTGDQREVGLSGTSAFNSVYKKVENTVLP 1500
DB 1441 RKXDSGVQSSHFLOGAKKNLSLALTLEMTGDQREVGLSGTSAFNSVYKKVENTVLP 1500
QY 1501 KPDLPTSGKVELLPKVHIYQKDLFPTETSNQSPGHLDLVEGSLLOCTEGAIAKWEANRP 1560

DB 1501 KPDLPTSGKVELLPKVHIYQKDLFPTETSNQSPGHLDLVEGSLLOCTEGAIAKWEANRP 1560
QY 1561 GKVPFLAVATESAKTPSKLIDPLANDNHYGTQIPKEEWSQBSPEKTAFAKKDITLSL 1620
DB 1561 GKVPFLAVATESAKTPSKLIDPLANDNHYGTQIPKEEWSQBSPEKTAFAKKDITLSL 1620
QY 1621 NACESHIAAINNEGQNKPEIEVTWAKQGTERTLCSQNPVLRKHQREIIRTTILQSDQEE 1680
DB 1621 NACESHIAAINNEGQNKPEIEVTWAKQGTERTLCSQNPVLRKHQREIIRTTILQSDQEE 1680
QY 1681 IDYDDTISVMKKEDFDIYDEDENQSPRSQKTRHYFIAAVERLWDYGMSSSPHVLRNR 1740
DB 1681 IDYDDTISVMKKEDFDIYDEDENQSPRSQKTRHYFIAAVERLWDYGMSSSPHVLRNR 1740
QY 1741 AQSGSVPOFKKVVQBEFTDGSFTQPLVRGELNHLGGLGPVIRAEVEDNMTVFRNOASR 1800
DB 1741 AQSGSVPOFKKVVQBEFTDGSFTQPLVRGELNHLGGLGPVIRAEVEDNMTVFRNOASR 1800
QY 1801 PYSFYSSLSIYEEEDQROGAEPKRFVKNPNETKITVFWKVOHHMAPTKDEDFCKAWAYFSDV 1860
DB 1801 PYSFYSSLSIYEEEDQROGAEPKRFVKNPNETKITVFWKVOHHMAPTKDEDFCKAWAYFSDV 1860
QY 1861 DLEKDVHSGLIGPLLCHTNTLPAHGRQVTVQEPALFFTFIDETKSWYFTENNERCRA 1920
DB 1861 DLEKDVHSGLIGPLLCHTNTLPAHGRQVTVQEPALFFTFIDETKSWYFTENNERCRA 1920
QY 1921 PCNTQMBDPTFKENYRFAHNGYIMDTLPGLVMAQDQRIWYLLSMGSENHSHIHPSGH 1980
DB 1921 PCNTQMBDPTFKENYRFAHNGYIMDTLPGLVMAQDQRIWYLLSMGSENHSHIHPSGH 1980
QY 1981 VFTVRKKEEYKMALYNLYPGVFETVEMLPKAGIRWVECLIGEHLHAGMSTLFLVYSNKC 2040
DB 1981 VFTVRKKEEYKMALYNLYPGVFETVEMLPKAGIRWVECLIGEHLHAGMSTLFLVYSNKC 2040
QY 2041 QTPGLMASGHIRDPQITASGOYGOWAPKRLARLHYSGSINAWSTKEPSPWIKVDLLAPMII 2100
DB 2041 QTPGLMASGHIRDPQITASGOYGOWAPKRLARLHYSGSINAWSTKEPSPWIKVDLLAPMII 2100
QY 2101 HGIKTQAROKFSSLYISQFIIMYSLDGKKWQTVRGNSTGTLVFFGKVDSSGKHNIFN 2160
DB 2101 HGIKTQAROKFSSLYISQFIIMYSLDGKKWQTVRGNSTGTLVFFGKVDSSGKHNIFN 2160
QY 2161 PPIIARVIRLEPHTHYSTIRLMEIMLGCDLNSCMPLGESKASISDAQITASSYFTNMFA 2220
DB 2161 PPIIARVIRLEPHTHYSTIRLMEIMLGCDLNSCMPLGESKASISDAQITASSYFTNMFA 2220
QY 2221 TWSPSKARLHLQGRSNARWPOVNNPKEWLOVDFOKTKMVTGVTTCQVKSLLTSMVVKFPL 2280
DB 2221 TWSPSKARLHLQGRSNARWPOVNNPKEWLOVDFOKTKMVTGVTTCQVKSLLTSMVVKFPL 2280
QY 2281 ISSSDQGHQWTLFFQNGKVKVFOGNDQSFTPVANSLDPPLLTRVLRHPQSWHQAIALRM 2340
DB 2281 ISSSDQGHQWTLFFQNGKVKVFOGNDQSFTPVANSLDPPLLTRVLRHPQSWHQAIALRM 2340
QY 2341 EYLGCEAQDLY 2351
DB 2341 EYLGCEAQDLY 2351
RESULT 10
AAW11445
ID AAW11445 standard; protein; 2351 AA.
XX
AC AAW11445;
XX
DT 20-NOV-1997 (first entry)
XX
DE Active Factor VIII:C analogue L1643X.
KW Factor VIII:C; analogue; glycoprotein; blood coagulation cascade;
KW fibrinogen; fibrin clot; haemostasis; haemophilia A; bleeding diathesis;
KW plasma protease; thrombin; immunogen; antibody; haemophilic; therapy;
KW

KW	proteolytic cleavage.	QY	121	GVSVKASGEAEYDDDOTSQREKEDDKVFPQGSHTYVWQVLKENGPMASDPLCLTYSYLSH	180
XX		Db	121	GVSVKASGEAEYDDDOTSQREKEDDKVFPQGSHTYVWQVLKENGPMASDPLCLTYSYLSH	180
OS	Homo sapiens.	QY	181	VDLVKOLNSGLIGALLVCREGSLAKETQTLHKFILLFAVFDGKSWHSTKNSLMODRD	240
XX	Synthetic.	Db	181	VDLVKOLNSGLIGALLVCREGSLAKETQTLHKFILLFAVFDGKSWHSTKNSLMODRD	240
XX	Key	QY	241	AASARAWPKXHTVNGYVNRSLPGLIGCHRXSVYWHVIGMGTTPVHIFLEGHTFLVRNH	300
FT	Peptide	Db	241	AASARAWPKXHTVNGYVNRSLPGLIGCHRXSVYWHVIGMGTTPVHIFLEGHTFLVRNH	300
FT	Protein	QY	301	QOASLEISPTITFLPAQTLMLMDLGOFLSCSISSHQHDMGEAYVKVDSCPPEPQURMKNE	360
FT	Region	Db	301	QOASLEISPTITFLPAQTLMLMDLGOFLSCSISSHQHDMGEAYVKVDSCPPEPQURMKNE	360
FT	Domain	QY	361	EAEYDDDLTDSMDVYRFDDDNPSFIQIRSAVKHCKPTWVHIAAEEDDWDVAPLVLA	420
FT	Modified-site	Db	361	EAEYDDDLTDSMDVYRFDDDNPSFIQIRSAVKHCKPTWVHIAAEEDDWDVAPLVLA	420
FT	Region	QY	421	PDRSYKSOYLNGPQIRIGRYKVRPMAVTDFTKTREAIQHESGILGPLLYGEVGDTL	480
FT		Db	421	PDRSYKSOYLNGPQIRIGRYKVRPMAVTDFTKTREAIQHESGILGPLLYGEVGDTL	480
XX	09-JUL-1996; 95WO-US011444.	QY	481	LIIFKQASRPYNIYPHGITDVRPLYSRRLPKGVKHLKDPFIPGFIKFKWTVVEDGP	540
XX	11-JUL-1995; 95US-0001025P.	Db	481	LIIFKQASRPYNIYPHGITDVRPLYSRRLPKGVKHLKDPFIPGFIKFKWTVVEDGP	540
XX	(CHIR) CHIRON CORP.	QY	541	TKSDPRCLTRYSSFNMRDLASGLIGPLIICYKESVDORGNQIMSDKRVILFSVFDE	600
XX	Hung DT, Cohen FE, Innis M;	Db	541	TKSDPRCLTRYSSFNMRDLASGLIGPLIICYKESVDORGNQIMSDKRVILFSVFDE	600
XX	WPI; 1997-119050/11.	QY	601	NRSWLTENIORFLPNPAGVQLEDPEQASNMHSINGYVDSLSQLSVCLHEVAYWILS	660
XX	Factor VIII:C analog modified adjacent to a non-activating Arg residue -	Db	601	NRSWLTENIORFLPNPAGVQLEDPEQASNMHSINGYVDSLSQLSVCLHEVAYWILS	660
XX	used in the treatment of haemophiliacs, by improvement of haemostasis.	QY	661	IGAOTDFLSVFPFGYTFKHKVVYEDTLTLPFPGSETVPMENGLWILGCHNSDFNRNG	720
XX	Claim 32; Page: 90pp; English.	Db	661	IGAOTDFLSVFPFGYTFKHKVVYEDTLTLPFPGSETVPMENGLWILGCHNSDFNRNG	720
XX	AAW11330-W11472 represent active Factor VIII:C analogues of the	QY	721	MTALKVSSCKNTGDDYEDSVEDISAYLLSKNNAIEPRSFQNSRHPSTKQKNATTI	780
XX	invention. These sequences were created by mutating the wild type Factor	Db	721	MTALKVSSCKNTGDDYEDSVEDISAYLLSKNNAIEPRSFQNSRHPSTKQKNATTI	780
XX	VIII:C coding sequence (see AWF1357) using mutagenic primers. The	QY	781	PENDIEKTDPAFHRTMPKIQNVSSSDLLMLLRQSPTHGLSLSDLOEAKYETFSDDPS	840
XX	analogues comprise a native Factor VIII:C polypeptide modified at a site	Db	781	PENDIEKTDPAFHRTMPKIQNVSSSDLLMLLRQSPTHGLSLSDLOEAKYETFSDDPS	840
XX	adjacent to a non-activating Arg residue so that a Arg-Pro or Pro-Arg	QY	841	PGAIDSNNSLSEMTFRPQLHHSQDMVTPRSGLOLRNEKLGTATAATELKLDFKVSST	900
XX	dipeptide is created. Factor VIII:C is a large glycoprotein that	Db	841	PGAIDSNNSLSEMTFRPQLHHSQDMVTPRSGLOLRNEKLGTATAATELKLDFKVSST	900
XX	participates in the blood coagulation cascade that ultimately converts	QY	901	NNLSLSTIPSDNLAAGTDNTSLGPPSPMVHYDSQDITTLFGKSSPLTEGGPLSLSEE	960
XX	soluble fibrinogen to insoluble fibrin clot effecting haemostasis. A	Db	901	NNLSLSTIPSDNLAAGTDNTSLGPPSPMVHYDSQDITTLFGKSSPLTEGGPLSLSEE	960
XX	deficiency in Factor VIII:C is responsible for haemophilia A, which is an	QY	961	NNDSKLLSGLMNSQESSGWKNVSTESGRIFPKGKRAHGPALLTKDNALFKVSI SLLKTN	1020
XX	X-chromosome-linked inherited bleeding diathesis. Factor VIII:C is	Db	961	NNDSKLLSGLMNSQESSGWKNVSTESGRIFPKGKRAHGPALLTKDNALFKVSI SLLKTN	1020
XX	activated by plasma proteases, such as thrombin. During activation the	QY	1021	KTSNNSATNRKTHIDGPSLLIENSPSVWQNLIESDTEFKKVTPLIHDRMLMDKNATLRL	1080
XX	mature polypeptide is cleaved to generate heavy and light chain fragments	Db	1021	KTSNNSATNRKTHIDGPSLLIENSPSVWQNLIESDTEFKKVTPLIHDRMLMDKNATLRL	1080
XX	that are further cleaved. Complexes of two or more of the analogues,	QY	1081	NHMSNKTTSKKNMNVQKKGPFPDQNPDMSPFKMLFLPESARWQRTQKKNLSNG	1140
XX	nucleic acids and vectors encoding them may be used alone or in	Db	1081	NHMSNKTTSKKNMNVQKKGPFPDQNPDMSPFKMLFLPESARWQRTQKKNLSNG	1140
XX	conjunction with each other, for the prevention or treatment of active	QY	1141	QGPSKQLVSLGPEKSVBQGNFLSKNVKVGKGFQDKVGLKEMVFPSSRNLFITNLDN	1200
XX	Factor VIII:C deficiency in a mammal. The analogues may be used as	Db	1141	QGPSKQLVSLGPEKSVBQGNFLSKNVKVGKGFQDKVGLKEMVFPSSRNLFITNLDN	1200
XX	immunogens to raise antibodies, and in the treatment of haemophiliacs, by	QY	1201	LEHNNYTHQEKILQBEIEKKEKTLIQENTVVLPOIH*VTGTGKNFMKNLFLILSTEQNVGSYD	1260
XX	improvement of haemostasis. The analogues are resistant to proteolytic	Db	1201	LEHNNYTHQEKILQBEIEKKEKTLIQENTVVLPOIH*VTGTGKNFMKNLFLILSTEQNVGSYD	1260
XX	cleavage and display increased plasma half-life. They may be administered	QY			
XX	at lower dosages and by different modes of administration	Db			
XX	Sequence 2351 AA;	QY			
SQ		Db			
	Query Match 99.9%; Score 12405; DB 2; Length 2351;				
	Best Local Similarity 99.9%; Pred. No. 0;				
	Matches 2349; Conservative 0; Mismatches 2; Indels 0; Gaps 0;				
QY	1 MQEISTCPFLCLLRFCFSATRYILGAVELSDWYMQSDLGELPVDARPPVPKSPFPN	60			
Db	1 MQEISTCPFLCLLRFCFSATRYILGAVELSDWYMQSDLGELPVDARPPVPKSPFPN	60			
QY	61 TSVVYKKTUFVEVTDHLFNIAKRPMPWMLLGPTIOAEVVDTVVITLKNMASHPVSLHAV	120			
Db	61 TSVVYKKTUFVEVTDHLFNIAKRPMPWMLLGPTIOAEVVDTVVITLKNMASHPVSLHAV	120			

Db 1201 LHENNTNQEKKEEIEKKEETLQSNVVLQIHTVGTGKFMKNLFLSLTRQNVESYD 1260
Qy 1261 GAYAPVLQDFSLANDSNRTKHTAHFSPKSGEENLEGLGNQTOIIVKCYACTRISPT 1320
Db 1261 GAYAPVLQDFSLANDSNRTKHTAHFSPKSGEENLEGLGNQTOIIVKCYACTRISPT 1320
Qy 1321 SQQNFTQSRKALQFRLPLEEETLEKRIIVDDTSTQWSKNMKHLEPSTLTQIDYNEKE 1380
Db 1321 SQQNFTQSRKALQFRLPLEEETLEKRIIVDDTSTQWSKNMKHLEPSTLTQIDYNEKE 1380
Qy 1381 KGALTQSPSLCLTRSHSIPQANESPLPIAKVSPSPSIRPIYLRVLFPQDNSSHLPAASY 1440
Db 1381 KGALTQSPSLCLTRSHSIPQANESPLPIAKVSPSPSIRPIYLRVLFPQDNSSHLPAASY 1440
Qy 1441 PKDSGVQBSSEHFTQGAKKNNLSAILTLEMGGQREVSGISGTATNSVTYKKKVENTVLP 1500
Db 1441 PKDSGVQBSSEHFTQGAKKNNLSAILTLEMGGQREVSGISGTATNSVTYKKKVENTVLP 1500
Qy 1501 KPDLPKTSKVELLPKVHIYOKDLPPTTSNGSPGHLDLVESLLQGTGEGALKWNEANRP 1560
Db 1501 KPDLPKTSKVELLPKVHIYOKDLPPTTSNGSPGHLDLVESLLQGTGEGALKWNEANRP 1560
Qy 1561 GKVPFLRVATSSAKTPSKLADPLAWDNHYGTQIPKEEWKQSEKSPKTAFFKKOTILSL 1620
Db 1561 GKVPFLRVATSSAKTPSKLADPLAWDNHYGTQIPKEEWKQSEKSPKTAFFKKOTILSL 1620
Qy 1621 NACSNHAIAINNEGONKPBIEVWAKQGRTERLCSQNPVVLKXHQREITRITLOSDEE 1680
Db 1621 NACSNHAIAINNEGONKPBIEVWAKQGRTERLCSQNPVVLKXHQREITRITLOSDEE 1680
Qy 1681 IDYDDTISVEMKKEDFDIYDEENOSPSFOKTRHYFIAAVERLWDMGSSPHVLNRP 1740
Db 1681 IDYDDTISVEMKKEDFDIYDEENOSPSFOKTRHYFIAAVERLWDMGSSPHVLNRP 1740
Qy 1741 AQSGSVPOFKVVFQFTDGSFTQPLRGELNEHLGGLGPIYRAEVENIMVTPFNQASR 1800
Db 1741 AQSGSVPOFKVVFQFTDGSFTQPLRGELNEHLGGLGPIYRAEVENIMVTPFNQASR 1800
Qy 1801 PYSFYSSLI SYEEDORQAEPRKFNKETHYFWKHQHMAPTDEDFCKAWAYFSDV 1860
Db 1801 PYSFYSSLI SYEEDORQAEPRKFNKETHYFWKHQHMAPTDEDFCKAWAYFSDV 1860
Qy 1861 DLEKDVHSLGLPLVCHTNTLNPAGHQVTVQEEFALPFTTFDETKSWYFTENMBNCR 1920
Db 1861 DLEKDVHSLGLPLVCHTNTLNPAGHQVTVQEEFALPFTTFDETKSWYFTENMBNCR 1920
Qy 1921 PCNIQMEDPTPKENVRFAHNGYIMDTLPGLVNAQDQRIWYLLSMGNSNENIHSIFSGH 1980
Db 1921 PCNIQMEDPTPKENVRFAHNGYIMDTLPGLVNAQDQRIWYLLSMGNSNENIHSIFSGH 1980
Qy 1981 VPTVRKKEEYKVALYNLYPGVETVEMLPKAGIWEVECLIGEHLHAGVSTLFLVYSNKC 2040
Db 1981 VPTVRKKEEYKVALYNLYPGVETVEMLPKAGIWEVECLIGEHLHAGVSTLFLVYSNKC 2040
Qy 2041 QTPGLWASGHIRDFOITASQYQGWAPKLARLHYSGSINAWSTKEPFSWKVDLLAPMII 2100
Db 2041 QTPGLWASGHIRDFOITASQYQGWAPKLARLHYSGSINAWSTKEPFSWKVDLLAPMII 2100
Qy 2101 HGITQGARQKSSYIYISOFIIMYSLDGKKWQTYRGNSTGTLMWFFGNVDSSGIRKHNIPN 2160
Db 2101 HGITQGARQKSSYIYISOFIIMYSLDGKKWQTYRGNSTGTLMWFFGNVDSSGIRKHNIPN 2160
Qy 2161 PPIIARYIRLHPHTHSIRSLRMEMLGCDLNSCPLGMSKAISSAQITASSYFNNMFA 2220
Db 2161 PPIIARYIRLHPHTHSIRSLRMEMLGCDLNSCPLGMSKAISSAQITASSYFNNMFA 2220
Qy 2221 TWSPSKARLHLOGRSNAPVQVNNPKELQVDFQKTMKVTGTTQGVKSLLSMYVKEFL 2280
Db 2221 TWSPSKARLHLOGRSNAPVQVNNPKELQVDFQKTMKVTGTTQGVKSLLSMYVKEFL 2280
Qy 2281 ISSSQDGHQWTLFFQNGKVKVFGQNDSTFPVNSLDPPLLTRYLRHQPQSWWHQIALRM 2340
Db 2281 ISSSQDGHQWTLFFQNGKVKVFGQNDSTFPVNSLDPPLLTRYLRHQPQSWWHQIALRM 2340

Db 2281 ISSSQDGHQWTLFFQNGKVKVFGQNDSTFPVNSLDPPLLTRYLRHQPQSWWHQIALRM 2340
Qy 2341 EVLGCERQDDLY 2351
Db 2341 EVLGCERQDDLY 2351
RESULT 11
AAW11398
ID AAW11398 standard; protein; 2351 AA.
XX AAW11398;
AC AAW11398;
DT 18-NOV-1997 (first entry)
XX Active Factor VIII:C analogue, delta 746, + residue 746 insertion.
DE Factor VII:C; analogue; glycoprotein; blood coagulation cascade;
KW fibrinogen; fibrin clot; haemostasis; haemophilia A; bleeding diathesis;
KW plasma protease; thrombin; immunogen; antibody; haemophiliac; therapy;
KW proteolytic cleavage.
XX Homo sapiens.
OS Synthetic.
XX
PH Key Location/Qualifiers
FT Peptide 1..19
FT Protein /note= "signal peptide"
FT /note= "mature Factor VIII:C"
FT Region 20..1667
FT /note= "heavy chain fragment"
FT Domain 760..1667
FT /note= "B domain"
FT Misc-difference 764..765
FT /note= "site of 1 residue deletion"
FT Modified-site 765
FT /note= "inserted residue"
FT Region 1668..2350
FT /note= "light chain fragment"
XX WC9703195-A1.
XX
PD 30-JAN-1997.
XX
PF 09-JUL-1996; 96WO-US011444.
XX
PR 11-JUL-1995; 95US-0001025P.
XX
PA (CHIR) CHIRON CORP.
XX
PI Hung DT, Cohen FE, Innis M;
XX
DR WPI; 1997-119050/11.
XX
PT Factor VIII:C analog modified adjacent to a non-activating Arg residue -
XX used in the treatment of haemophiliacs, by improvement of haemostasis.
XX
PS Claim 23; Page; 90pp; English.
XX
CC AAW11330-W11472 represent active Factor VIII:C analogues of the
CC invention. These sequences were created by mutating the wild type Factor
CC VIII:C coding sequence (see AAT51357) using mutagenic primers. The
CC analogues comprise a native Factor VIII:C polypeptide modified at a site
CC adjacent to a non-activating Arg residue so that a Arg-Pro or Pro-Arg
CC dipeptide is created. Factor VIII:C is a large glycoprotein that
CC participates in the blood coagulation cascade that ultimately converts
CC soluble fibrinogen to insoluble fibrin clot, effecting haemostasis. A
CC deficiency in Factor VIII:C is responsible for haemophilia A, which is an
CC X-chromosome-linked inherited bleeding diathesis. Factor VIII:C is
CC activated by plasma proteases, such as thrombin. During activation the
CC mature polypeptide is cleaved to generate heavy and light chain fragments
CC that are further cleaved. Complexes of two or more of the analogues,

CC nucleic acids and vectors encoding them may be used alone or in
 CC conjunction with each other, for the prevention or treatment of active
 CC Factor VIII:C deficiency in a mammal. The analogues may be used as
 CC immunogens to raise antibodies, and in the treatment of haemophilias, by
 CC improvement of haemostasis. The analogues are resistant to proteolytic
 CC cleavage and display increased plasma half-life. They may be administered
 CC at lower dosages and by different modes of administration
 XX

Sequence 2351 AA;

Query Match	99.9%;	Score 12405;	DB 2;	Length 2351;
Best Local Similarity	99.9%;	Pred. No. 0;		
Matches 2349;	Conservative	0;	Mismatches 2;	Indels 0;
Gaps	0;			

QY	1	MOELSTCFPLCLLPCFSA	RYVILGAVELSDYMQSD	LGELPVDARPPVPKSFPPN	60
DB	1	MOELSTCFPLCLLPCFSA	RYVILGAVELSDYMQSD	LGELPVDARPPVPKSFPPN	60
QY	61	TSVYVKTLFVEFTD	LFNIAKPRPPMGLLPT	CAEYDVTWITLKNASHPVSLHAV	120
DB	61	TSVYVKTLFVEFTD	LFNIAKPRPPMGLLPT	CAEYDVTWITLKNASHPVSLHAV	120
QY	121	GVSYWKASGAEYDDOTS	QREKEDDKVPFGGSHTYVW	YLVKENGPMASDPLCLTYSYLH	180
DB	121	GVSYWKASGAEYDDOTS	QREKEDDKVPFGGSHTYVW	YLVKENGPMASDPLCLTYSYLH	180
QY	181	VDLVXDLNSGLIGALLV	CEGSLAKKTQTLHKFILL	FAVDEGKSWHSETXNSIMOOD	240
DB	181	VDLVXDLNSGLIGALLV	CEGSLAKKTQTLHKFILL	FAVDEGKSWHSETXNSIMOOD	240
QY	241	AASARAWPKMHTVNGV	NRSLPGLIGCHRKSVVHV	IGMGTTPVHSHIFLEGHTFLVRNH	300
DB	241	AASARAWPKMHTVNGV	NRSLPGLIGCHRKSVVHV	IGMGTTPVHSHIFLEGHTFLVRNH	300
QY	301	ROASLEISITELTAQTL	MDLQGFLLSCHISHOH	DGMEAYVKVDCSPEEPOLRMKNE	360
DB	301	ROASLEISITELTAQTL	MDLQGFLLSCHISHOH	DGMEAYVKVDCSPEEPOLRMKNE	360
QY	361	EAEYDDDLTDSMDVVR	DDNSPSFIOIRSVAKX	HPKTHVHYIAAEBEDWDYAPLVLA	420
DB	361	EAEYDDDLTDSMDVVR	DDNSPSFIOIRSVAKX	HPKTHVHYIAAEBEDWDYAPLVLA	420
QY	421	PDRSYKSVLNGPQRI	GRKYKVRPMAYDTDF	FTREAIQHESGILGPLLYCEVGDIL	480
DB	421	PDRSYKSVLNGPQRI	GRKYKVRPMAYDTDF	FTREAIQHESGILGPLLYCEVGDIL	480
QY	481	LIIFKQASRPYNIYPH	GITDVEPLYSRRLPK	GVKHLKOPFPLPGHIFYKWTVTVEDGP	540
DB	481	LIIFKQASRPYNIYPH	GITDVEPLYSRRLPK	GVKHLKOPFPLPGHIFYKWTVTVEDGP	540
QY	541	TKSDPECLTRYSSFV	NMERDLASGLIGPLI	CYKESVDQGNQIYSDXENVILPSVEDE	600
DB	541	TKSDPECLTRYSSFV	NMERDLASGLIGPLI	CYKESVDQGNQIYSDXENVILPSVEDE	600
QY	601	NRSWYLTEIQRLEP	NAPAGVQLEDPEPQ	ASNIHMSINGYVFDLSQLSVCLHEVAYWYLS	660
DB	601	NRSWYLTEIQRLEP	NAPAGVQLEDPEPQ	ASNIHMSINGYVFDLSQLSVCLHEVAYWYLS	660
QY	661	IGAQTDFLSVFPFGY	TPKHNVEDTLTPP	PGSETVFMSENPGWLILCHNSDFNRG	720
DB	661	IGAQTDFLSVFPFGY	TPKHNVEDTLTPP	PGSETVFMSENPGWLILCHNSDFNRG	720
QY	721	MTALLKVSSCDKNTG	YIYEDSYSDISAYLS	SKNAIEPRFSQNSRHPSTRQGFNATTI	780
DB	721	MTALLKVSSCDKNTG	YIYEDSYSDISAYLS	SKNAIEPRFSQNSRHPSTRQGFNATTI	780
QY	781	PENDIEKTDWFAHRT	PMPKIQNVSSDMLL	MLLRQSPTPHGLSLDLQEAKEYTFSDDPS	840
DB	781	PENDIEKTDWFAHRT	PMPKIQNVSSDMLL	MLLRQSPTPHGLSLDLQEAKEYTFSDDPS	840
QY	841	PGAIDSNNSLSSEMT	HPQHLHSGDMVFTD	ASGLQLRNEKLGTTAATLKKLDKFKVSS	900
DB	841	PGAIDSNNSLSSEMT	HPQHLHSGDMVFTD	ASGLQLRNEKLGTTAATLKKLDKFKVSS	900

QY	901	SNNLISIPSDNLAAGT	NTSSLGPPSPMPEHYD	SOLDTTLFGKKSSPLTESGGPLSSEE	960
DB	901	SNNLISIPSDNLAAGT	NTSSLGPPSPMPEHYD	SOLDTTLFGKKSSPLTESGGPLSSEE	960
QY	961	NNDKSLFESGLMNSQ	ESSWGKNVSTESGR	LFGKRAHGPALLTKDNALFKVSI	1020
DB	961	NNDKSLFESGLMNSQ	ESSWGKNVSTESGR	LFGKRAHGPALLTKDNALFKVSI	1020
QY	1021	KTSNNSATNRKTHID	GPSELLIENSFVQNI	LESDETEPKKVTPLIHDRMLKDKATARL	1080
DB	1021	KTSNNSATNRKTHID	GPSELLIENSFVQNI	LESDETEPKKVTPLIHDRMLKDKATARL	1080
QY	1081	NHMSNKTSSKNM	VMQKKEGPIPPDA	QNPDMPFFKMLFPESARWQRTHGKNSLNSG	1140
DB	1081	NHMSNKTSSKNM	VMQKKEGPIPPDA	QNPDMPFFKMLFPESARWQRTHGKNSLNSG	1140
QY	1141	QGPSKQVLVSLGP	BEKVEGQNFLEKKN	VVVGKGEFTKDVGLKEMVFPSSNLFUTNLDN	1200
DB	1141	QGPSKQVLVSLGP	BEKVEGQNFLEKKN	VVVGKGEFTKDVGLKEMVFPSSNLFUTNLDN	1200
QY	1201	LHENNTHNQEKKI	QBEIEKKE	TLOENVVLQIHTVTCTKNPMKNLFLLS	1260
DB	1201	LHENNTHNQEKKI	QBEIEKKE	TLOENVVLQIHTVTCTKNPMKNLFLLS	1260
QY	1261	GAYAPVLODFRSL	NDSTNRKTKHTAH	FKSKGEEENLEGLNQTKOIVEKYACTTRISPT	1320
DB	1261	GAYAPVLODFRSL	NDSTNRKTKHTAH	FKSKGEEENLEGLNQTKOIVEKYACTTRISPT	1320
QY	1321	SOQNFVTRQSRK	ALQKPLP	LEETELEXRIIVDDTSTOWSKMKHLTSTLT	1380
DB	1321	SOQNFVTRQSRK	ALQKPLP	LEETELEXRIIVDDTSTOWSKMKHLTSTLT	1380
QY	1381	KGAIITQSP	SDCLTSHSIT	POANRSPPLPIAKVSSPPSIRPIYLT	1440
DB	1381	KGAIITQSP	SDCLTSHSIT	POANRSPPLPIAKVSSPPSIRPIYLT	1440
QY	1441	RKDGSGV	QESSHFLGAKKNN	LSLAILTLEMTGDQREVGLSTSATNSVYTKKVENTVLP	1500
DB	1441	RKDGSGV	QESSHFLGAKKNN	LSLAILTLEMTGDQREVGLSTSATNSVYTKKVENTVLP	1500
QY	1501	KPDLPTSGKV	ELLPKVHIYQKDL	PPTETNSGSPHOLDVREGSLQGTGEGALIKWNEANRP	1560
DB	1501	KPDLPTSGKV	ELLPKVHIYQKDL	PPTETNSGSPHOLDVREGSLQGTGEGALIKWNEANRP	1560
QY	1561	GKVPF	RYATESAKTPSK	LLDPLAWDNHYGTQIPKEWKSOEKSPEKTA	1620
DB	1561	GKVPF	RYATESAKTPSK	LLDPLAWDNHYGTQIPKEWKSOEKSPEKTA	1620
QY	1621	NACESNHAIAA	INEGONKEPEI	EVTWAKQGRTERLCSQNPVLKRHQRETRITTLQSDQEE	1680
DB	1621	NACESNHAIAA	INEGONKEPEI	EVTWAKQGRTERLCSQNPVLKRHQRETRITTLQSDQEE	1680
QY	1681	IDYDDTIS	YEMKKEPDIY	DEBENOSPRSQKTRHYFIAAVERLWDYGMSSPHVLRNR	1740
DB	1681	IDYDDTIS	YEMKKEPDIY	DEBENOSPRSQKTRHYFIAAVERLWDYGMSSPHVLRNR	1740
QY	1741	AQSGSV	QPKVYVQEF	TGSGFTQPLYRGELNEHLGLGPIYRAEVEDNIMWTFRNOASR	1800
DB	1741	AQSGSV	QPKVYVQEF	TGSGFTQPLYRGELNEHLGLGPIYRAEVEDNIMWTFRNOASR	1800
QY	1801	PYSFYSSLIS	YEBEDQ	RQGAEPKPNFVKPNETKITYPWKQVHMAPTKDEPDKAWAYFSDV	1860
DB	1801	PYSFYSSLIS	YEBEDQ	RQGAEPKPNFVKPNETKITYPWKQVHMAPTKDEPDKAWAYFSDV	1860
QY	1861	DLEKQVHSG	LIGLIP	LCHVTINTLNPAGROVTVQEFALFTIIFDETKSWYFTENMEENCA	1920
DB	1861	DLEKQVHSG	LIGLIP	LCHVTINTLNPAGROVTVQEFALFTIIFDETKSWYFTENMEENCA	1920
QY	1921	PCNIQMED	TFKENYEFHAI	NGYIMDTLPGVMAQDQRIWRVYLLSYGNSNENHSHIFSGH	1980
DB	1921	PCNIQMED	TFKENYEFHAI	NGYIMDTLPGVMAQDQRIWRVYLLSYGNSNENHSHIFSGH	1980

QY 1981 VFTVRKKEEYKMALYNLYPGVFETVEMLPKAGIWRVECLIGELHAGMSTFLVYSNKC 2040
 DB 1981 VFTVRKKEEYKMALYNLYPGVFETVEMLPKAGIWRVECLIGELHAGMSTFLVYSNKC 2040
 QY 2041 QTPIGMASGHIRDFOITASGOYGOWAPKLAARLHYSGSINAWSTKEPFSWIKVDLLAPMII 2100
 DB 2041 QTPIGMASGHIRDFOITASGOYGOWAPKLAARLHYSGSINAWSTKEPFSWIKVDLLAPMII 2100
 QY 2101 HGKIQARQKFSLSLYISQFLIMYSLDGKKWQTYRGNSTGTLMVFFGNVDSSGKHNEN 2160
 DB 2101 HGKIQARQKFSLSLYISQFLIMYSLDGKKWQTYRGNSTGTLMVFFGNVDSSGKHNEN 2160
 QY 2161 PPIIARVIRLHPTYSIRSTRIMELMGCDLNSCMPGMSKALSDAQIITASSYFTNNEFA 2220
 DB 2161 PPIIARVIRLHPTYSIRSTRIMELMGCDLNSCMPGMSKALSDAQIITASSYFTNNEFA 2220
 QY 2221 TWSPSKARLHLQGRSNAWRPQNNPKWLQVDFQKTMKVTVTTQGVKSLTSMYVKEFL 2280
 DB 2221 TWSPSKARLHLQGRSNAWRPQNNPKWLQVDFQKTMKVTVTTQGVKSLTSMYVKEFL 2280
 QY 2281 ISSSQDGHQWTLRFQNGKVKVFOGNQDSFTPVNSLOPPLCTRYLRTHPOSWVHQIALRM 2340
 DB 2281 ISSSQDGHQWTLRFQNGKVKVFOGNQDSFTPVNSLOPPLCTRYLRTHPOSWVHQIALRM 2340
 QY 2341 EVLGCEAQDLY 2351
 DB 2341 EVLGCEAQDLY 2351

RESULT 12

AAW11425

ID AAW11425 standard; protein; 2351 AA.

XX AC AAW11425;

XX DT 20-NOV-1997 (first entry)

DE DE Active Factor VIII:C analogue T1308X.

XX KW Factor VIII:C; analogue; glycoprotein; blood coagulation cascade;
 KW fibrinogen; fibrin clot; haemostasis; haemophilia A; bleeding diathesis;
 KW plasma protease; thrombin; immunogen; antibody; haemophilic; therapy;
 XX proteolytic cleavage.

XX OS Homo sapiens.

XX OS Synthetic.

XX FH Location/Qualifiers
 FT Peptide 1..19
 FT /note= "signal peptide"
 FT Protein 20..2351
 FT /note= "mature Factor VIII:C"
 FT Region 20..1667
 FT /note= "heavy chain fragment"
 FT Domain 760..1667
 FT /note= "B domain"
 FT Modified-site 1327
 FT /label= Phe, Glu, Pro
 FT Region 1668..2350
 FT /note= "light chain fragment"

XX PN WO9703195-A1.

XX PD 30-JAN-1997.

XX PF 09-JUL-1996; 96WO-US011444.

XX PR 11-JUL-1995; 95US-0001025P.

XX PA (CHIR) CHIRON CORP.

XX PI Fung DT, Cohen FE, Innis M;

XX XX

DR WPI: 1997-119050/11.
 XX Factor VIII:C analog modified adjacent to a non-activating Arg residue -
 PT used in the treatment of haemophilias, by improvement of haemostasis.
 XX Claim 28; Page; 90pp; English.
 CC AAW11330-W11472 represent active Factor VIII:C analogues of the
 CC invention. These sequences were created by mutating the wild type Factor
 CC VIII:C coding sequence (see NAT51357) using mutagenic primers. The
 CC analogues comprise a native Factor VIII:C polypeptide modified at a site
 CC adjacent to a non-activating Arg residue so that a Arg-Bro or Pro-Arg
 CC dipeptide is created. Factor VIII:C is a large glycoprotein that
 CC participates in the blood coagulation cascade that ultimately converts
 CC soluble fibrinogen to insoluble fibrin clot, effecting haemostasis. A
 CC deficiency in Factor VIII:C is responsible for haemophilia A, which is an
 CC X-chromosome-linked inherited bleeding diathesis. Factor VIII:C is
 CC activated by plasma proteases, such as thrombin. During activation the
 CC mature polypeptide is cleaved to generate heavy and light chain fragments
 CC that are further cleaved. Complexes of two or more of the analogues,
 CC nucleic acids and vectors encoding them may be used alone or in
 CC conjunction with each other, for the prevention or treatment of active
 CC Factor VIII:C deficiency in a mammal. The analogues may be used as
 CC immunogens to raise antibodies, and in the treatment of haemophilias, by
 CC improvement of haemostasis. The analogues are resistant to proteolytic
 CC cleavage and display increased plasma half-life. They may be administered
 CC at lower dosages and by different modes of administration
 XX SQ Sequence 2351 AA;

Query Match 99.9%; Score 12405; DB 2; Length 2351;
 Best Local Similarity 99.9%; Pred. No. 0;
 Matches 2349; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 1 MQIELSTCFE-CLLRFCFSATRYVLGAVELSDWYMSDLGELPVDARFPRVPKSPFN 60
 DB 1 MQIELSTCFE-CLLRFCFSATRYVLGAVELSDWYMSDLGELPVDARFPRVPKSPFN 60
 QY 61 TSVVYKTLFVEFTDHLFNIAKPRPPMMGLLGPTIOAEVYDVTVTLLKNMASHPVSLHAV 120
 DB 61 TSVVYKTLFVEFTDHLFNIAKPRPPMMGLLGPTIOAEVYDVTVTLLKNMASHPVSLHAV 120
 QY 121 GVSVMKASEGAEYDDQTSQREKDDKVPFGGSHYTVVQVLKENGPMASDPLCLTYSYLSH 180
 DB 121 GVSVMKASEGAEYDDQTSQREKDDKVPFGGSHYTVVQVLKENGPMASDPLCLTYSYLSH 180
 QY 181 VDLVKDLNSGLIGALLVCRGSLAKEKTQTLHKPILLFAVDEGKSWHSETKNSLMQDRD 240
 DB 181 VDLVKDLNSGLIGALLVCRGSLAKEKTQTLHKPILLFAVDEGKSWHSETKNSLMQDRD 240
 QY 241 AASAPAWPKMHTVNGYVNRSLFGLIGCHRXKVYVHVIGMGTTPVHSIFLEGHTFLVRNH 300
 DB 241 AASAPAWPKMHTVNGYVNRSLFGLIGCHRXKVYVHVIGMGTTPVHSIFLEGHTFLVRNH 300
 QY 301 ROASLEISPIITFLTAQTLMDLQGLLFSCHSSHQDGMAYKVYDSCPESEQLMKNE 360
 DB 301 ROASLEISPIITFLTAQTLMDLQGLLFSCHSSHQDGMAYKVYDSCPESEQLMKNE 360
 QY 361 EAEDYDDDLTDSEMDVVRFDNDNSPFIQIRSAVKKPKTWVHYIAAEEDWDYAPLVLA 420
 DB 361 EAEDYDDDLTDSEMDVVRFDNDNSPFIQIRSAVKKPKTWVHYIAAEEDWDYAPLVLA 420
 QY 421 PDDRSYKSOYLANGPQRIGRYKRVFMAYTDETFTKREAIQHSIGILGPLLYGVGDTL 480
 DB 421 PDDRSYKSOYLANGPQRIGRYKRVFMAYTDETFTKREAIQHSIGILGPLLYGVGDTL 480
 QY 481 LIIFKNQASRPYNTYPHGITDVRPLYSRLPKGVGHLKDFPILPGEI PKYKWTVTVEGDP 540
 DB 481 LIIFKNQASRPYNTYPHGITDVRPLYSRLPKGVGHLKDFPILPGEI PKYKWTVTVEGDP 540
 QY 541 TKSDPRCLTRYVSSFWNMERDLASGLIGPLLCYKESVDQRGNQIMSKRNVILSVFDE 600
 DB 541 TKSDPRCLTRYVSSFWNMERDLASGLIGPLLCYKESVDQRGNQIMSKRNVILSVFDE 600

QY 601 NRSYLTENTORFPLNACVOLPEBPQASNMHSINGVYFDSIQLSCLHEVAVWYLS 660
DB 601 NRSYLTENTORFPLNACVOLPEBPQASNMHSINGVYFDSIQLSCLHEVAVWYLS 660
QY 661 IGAGTDFLSVFFSGYTFKHKVYEDTTLTLPFSGETVFMSENPGMLILGCHNSDFNRG 720
DB 661 IGAGTDFLSVFFSGYTFKHKVYEDTTLTLPFSGETVFMSENPGMLILGCHNSDFNRG 720
QY 721 MTALLKVSCKNWDYEDSYEDISAYLLSKNNAIEPRSPQNSRHPSTKQKFNATTI 780
DB 721 MTALLKVSCKNWDYEDSYEDISAYLLSKNNAIEPRSPQNSRHPSTKQKFNATTI 780
QY 781 PENDIEKTDPMFAHRTMPKIQNVSSDLMMLLQSPHPHGLSLSDLOEAYEYFSDDP 840
DB 781 PENDIEKTDPMFAHRTMPKIQNVSSDLMMLLQSPHPHGLSLSDLOEAYEYFSDDP 840
QY 841 PGAIDSNLSGEMTHFRPQLHSGDMVFTPSGQLRLNEKLGTTAAATELKKLDFKYSST 900
DB 841 PGAIDSNLSGEMTHFRPQLHSGDMVFTPSGQLRLNEKLGTTAAATELKKLDFKYSST 900
QY 901 SNNLITIPSNLAAGTNTSSLPSPMPVYDSQDITLFGKSSPLTESGGPLSLEE 960
DB 901 SNNLITIPSNLAAGTNTSSLPSPMPVYDSQDITLFGKSSPLTESGGPLSLEE 960
QY 961 NNDKSLLESGLMNSOESSGKNVSTESGRFLFKGRAHPALLTKDNALFKVSTISLLKTN 1020
DB 961 NNDKSLLESGLMNSOESSGKNVSTESGRFLFKGRAHPALLTKDNALFKVSTISLLKTN 1020
QY 1021 KTSNNSATNRKTHIDGPSILLIENSESVWQNLLESDETEFKVTPLIHDMKMDKATALLR 1080
DB 1021 KTSNNSATNRKTHIDGPSILLIENSESVWQNLLESDETEFKVTPLIHDMKMDKATALLR 1080
QY 1081 NEMSKNTTSSKNMVMYQKKGPIPPDAQNPMSPFKMLFPESARWQRTGKNSNSG 1140
DB 1081 NEMSKNTTSSKNMVMYQKKGPIPPDAQNPMSPFKMLFPESARWQRTGKNSNSG 1140
QY 1141 QGPKQOLVSLGPEKSVGQNFPLSKNVVYVGKGFTKDVGKEMVFPSSNLFITNLDN 1200
DB 1141 QGPKQOLVSLGPEKSVGQNFPLSKNVVYVGKGFTKDVGKEMVFPSSNLFITNLDN 1200
QY 1201 LHENNTNHOEKIOEIEPKETLIGQNVVLPOIHTVTGTQNMKMLFLLSTRQNVGSYD 1260
DB 1201 LHENNTNHOEKIOEIEPKETLIGQNVVLPOIHTVTGTQNMKMLFLLSTRQNVGSYD 1260
QY 1261 GAYAPVLQDFSLNDSTNRKTKHTAHFSGKGEENLEGIGNTKOIVEKIACITRISPT 1320
DB 1261 GAYAPVLQDFSLNDSTNRKTKHTAHFSGKGEENLEGIGNTKOIVEKIACITRISPT 1320
QY 1321 SQCNFVTOQRKEALKQFRLPLETELEKRIIVDDTSTQMSKNMKHLTPSTLTQIDYNEKE 1380
DB 1321 SQCNFVTOQRKEALKQFRLPLETELEKRIIVDDTSTQMSKNMKHLTPSTLTQIDYNEKE 1380
QY 1381 KGAIQTQSPSLDCLTRSHSIPQANRPLPIAKVSSPSPRIPIYLTTRVLQDNDSHLPAA 1440
DB 1381 KGAIQTQSPSLDCLTRSHSIPQANRPLPIAKVSSPSPRIPIYLTTRVLQDNDSHLPAA 1440
QY 1441 RKKDSGVQSSHPFLOGAKNNLSALLTLEMTGDQREVGLSATSNTVYKKVNTVLP 1500
DB 1441 RKKDSGVQSSHPFLOGAKNNLSALLTLEMTGDQREVGLSATSNTVYKKVNTVLP 1500
QY 1501 KPDLPKTSKVELLPKVHTYQKDLFPETSTNSGPHGLDVEGSLQGTGEGAIKWEANRP 1560
DB 1501 KPDLPKTSKVELLPKVHTYQKDLFPETSTNSGPHGLDVEGSLQGTGEGAIKWEANRP 1560
QY 1561 GKVPFLRVATSSAKTPSKLPLDPLANDNHYGTQIPKEEWSQESKPEKTAFFKXDTILSL 1620
DB 1561 GKVPFLRVATSSAKTPSKLPLDPLANDNHYGTQIPKEEWSQESKPEKTAFFKXDTILSL 1620
QY 1621 NACESNHAIAANEGQNKPEIEVTWAKQRTSELCSQNPVLRKHOREITRITLQSDQEE 1680
DB 1621 NACESNHAIAANEGQNKPEIEVTWAKQRTSELCSQNPVLRKHOREITRITLQSDQEE 1680

QY 1681 IDYDDTISVENKKEDFDIYDEDENQSPRSFQKKTRHYFIAAVERLWDYKSSSPHVLRNR 1740
DB 1681 IDYDDTISVENKKEDFDIYDEDENQSPRSFQKKTRHYFIAAVERLWDYKSSSPHVLRNR 1740
QY 1741 AQSGSVPOFKKVPQBPFTDGSFTQPLYRGELNEHLGLLGYIRAEVEDNLKVTFRNQASR 1800
DB 1741 AQSGSVPOFKKVPQBPFTDGSFTQPLYRGELNEHLGLLGYIRAEVEDNLKVTFRNQASR 1800
QY 1801 PYSFYSSLIISVEDOROGABPRKXFKVPNETKHYFWKVOHEMAFTKDEPFCKAWAYFSDV 1860
DB 1801 PYSFYSSLIISVEDOROGABPRKXFKVPNETKHYFWKVOHEMAFTKDEPFCKAWAYFSDV 1860
QY 1861 DLEKDVHSGILGPLLVCHTNTLPAHGRQVTVQEFALFTTIFDSTKSWYFENMERNCR 1920
DB 1861 DLEKDVHSGILGPLLVCHTNTLPAHGRQVTVQEFALFTTIFDSTKSWYFENMERNCR 1920
QY 1921 PCNTQMEDPTFKENYRPHAINGYIMDTLPGLVMAQDQRIWYLLSMGSENENIHSIHFSGH 1980
DB 1921 PCNTQMEDPTFKENYRPHAINGYIMDTLPGLVMAQDQRIWYLLSMGSENENIHSIHFSGH 1980
QY 1981 VFTVRKKEBYKVALYNLYPGVFETVEMLPKSKAGIWRVECLIGEHLHAGMSTLFLVYSNKC 2040
DB 1981 VFTVRKKEBYKVALYNLYPGVFETVEMLPKSKAGIWRVECLIGEHLHAGMSTLFLVYSNKC 2040
QY 2041 QTPLGMASGHIRDFOITASGOYGOWAPKLARLHYSGSINAWSTKEPFSWIKVDLLAPMII 2100
DB 2041 QTPLGMASGHIRDFOITASGOYGOWAPKLARLHYSGSINAWSTKEPFSWIKVDLLAPMII 2100
QY 2101 HGIKTQAROKFSSLIYSOFTIMYSLDGKKWQTYRGNSTGTLMVFPFGNVDSGKHNIFN 2160
DB 2101 HGIKTQAROKFSSLIYSOFTIMYSLDGKKWQTYRGNSTGTLMVFPFGNVDSGKHNIFN 2160
QY 2161 PPIIARVIRLHPHTHYRSRSTLRMELMGCDLNSCMPLEGSKAISDAQITASSYFTNMFA 2220
DB 2161 PPIIARVIRLHPHTHYRSRSTLRMELMGCDLNSCMPLEGSKAISDAQITASSYFTNMFA 2220
QY 2221 TWSFSKARLHLQGRSNARWPOVNNPKWLQVDFPQKTMKVTVGTTQGVKSLTSMYKVEFL 2280
DB 2221 TWSFSKARLHLQGRSNARWPOVNNPKWLQVDFPQKTMKVTVGTTQGVKSLTSMYKVEFL 2280
QY 2281 ISSQDGHQWTLFPQNGKVKVFCQGNQDSFTPVVNSLDPPELLTRYLRHPQSWVHQIALRM 2340
DB 2281 ISSQDGHQWTLFPQNGKVKVFCQGNQDSFTPVVNSLDPPELLTRYLRHPQSWVHQIALRM 2340
QY 2341 EVLGCEAQDLY 2351
DB 2341 EVLGCEAQDLY 2351
RESULT 13
AAW11362
ID AAW11362 standard; protein; 2351 AA.
XX
AC AAW11362;
XX
DT 18-NOV-1997 (first entry)
XX
DE Active Factor VIII:C analogue L277X.
KW Factor VIII:C; analogue; glycoprotein; blood coagulation cascade;
KW fibrinogen; fibrin clot; haemostasis; haemophilia A; bleeding diathesis;
KW plasma protease; thrombin; immunogen; antibody; haemophilic; therapy;
XX proteolytic cleavage.
OS Homo sapiens.
OS Synthetic.
XX
FH Key
FT Peptide
FT Protein
FT Region
/note="signal peptide"
/note="2351"
/note="mature Factor VIII:C"
20.1667

FT /note= "heavy chain fragment"
 FT 236
 FT /label= Phe, Glu, Pro
 FT Domain 760..1667
 FT /note= "B domain"
 FT Region 1668..2350
 FT /note= "light chain fragment"
 XX
 DN W09703195-A1.
 XX
 XX 30-JAN-1997.
 XX
 XX 09-JUL-1996; 96WO-US011444.
 XX
 XX 11-JUL-1995; 95US-0001025P.
 XX
 XX (CHIR) CHIRON CORP.
 XX
 XX Hung DT, Cohen PE, Innis M;
 XX
 XX WPI; 1997-119050/11.
 DR
 XX Factor VIII:C analog modified adjacent to a non-activating Arg residue -
 XX used in the treatment of haemophilias, by improvement of haemostasis.
 PT
 XX
 XX Claim 14; Page; 90pp; English.
 XX
 CC AAW11330-W11472 represent active Factor VIII:C analogues of the
 CC invention. These sequences were created by mutating the wild type Factor
 CC VIII:C coding sequence (see AAT51357) using mutagenic primers. The
 CC analogues comprise a native Factor VIII:C polypeptide modified at a site
 CC adjacent to a non-activating Arg residue so that a Arg-Pro or Pro-Arg
 CC dipeptide is created. Factor VIII:C is a large glycoprotein that
 CC participates in the blood coagulation cascade that ultimately converts
 CC soluble fibrinogen to insoluble fibrin clot, effecting haemostasis. A
 CC deficiency in Factor VIII:C is responsible for haemophilia A, which is an
 CC x-chromosome-linked inherited bleeding diathesis. Factor VIII:C is
 CC activated by plasma proteases, such as thrombin. During activation the
 CC mature polypeptide is cleaved to generate heavy and light chain fragments
 CC that are further cleaved. Complexes of two or more of the analogues,
 CC nucleic acids and vectors encoding them may be used alone or in
 CC conjunction with each other, for the prevention or treatment of active
 CC Factor VIII:C deficiency in a mammal. The analogues may be used as
 CC immunogens to raise antibodies, and in the treatment of haemophilias, by
 CC improvement of haemostasis. The analogues are resistant to proteolytic
 CC cleavage and display increased plasma half-life. they may be administered
 CC at lower dosages and by different modes of administration
 XX
 SQ Sequence 2351 AA;
 Query Match 99.9%; Score 12405; DB 2; Length 2351;
 Best Local Similarity 99.9%; Pred. No. 0;
 Matches 2349; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 1 MQEILSTCFLLCLRFCSATRRYYLGAVALSWDYMQSDLGELPVDARFPVRPKSPFPN 60
 DB 1 MQEILSTCFLLCLRFCSATRRYYLGAVALSWDYMQSDLGELPVDARFPVRPKSPFPN 60
 QY 61 TSVVYKTLFVEFTDHLNIAKRPFPWGLLGPTTQABVYDTVITLKNMASHPVSLHAV 120
 DB 61 TSVVYKTLFVEFTDHLNIAKRPFPWGLLGPTTQABVYDTVITLKNMASHPVSLHAV 120
 QY 121 GVSVMKASEGABYDDQTSOREKEDDKVPFGSGHTYVWVLKENGPMASDPLCLTYSLSH 180
 DB 121 GVSVMKASEGABYDDQTSOREKEDDKVPFGSGHTYVWVLKENGPMASDPLCLTYSLSH 180
 QY 161 VDLVKDLSGLIGALLVCREGSLAKKXTQLHKFILLFAVDEGKSWSEYKNSLMQDRD 240
 DB 161 VDLVKDLSGLIGALLVCREGSLAKKXTQLHKFILLFAVDEGKSWSEYKNSLMQDRD 240
 QY 241 AASARAWPKMHTVNGVYVNRSLDPLIGCHRKSKSYWHVINGTTPPEVHSIFLEGHTFLVRNH 300
 DB 241 AASARAWPKMHTVNGVYVNRSLDPLIGCHRKSKSYWHVINGTTPPEVHSIFLEGHTFLVRNH 300

QY 301 ROASLEISPIITFLTAOTLLMDLGOPLISCHISSHOHDGMEAYVKVDSCPPEPQRLMKNNE 360
 DB 301 ROASLEISPIITFLTAOTLLMDLGOPLISCHISSHOHDGMEAYVKVDSCPPEPQRLMKNNE 360
 QY 361 EAEYDDEDLTDEMDVVRFDNDKSPSFIQIRSAKHKPTWVHYIAAEEEDWDYAPLVLA 420
 DB 361 EAEYDDEDLTDEMDVVRFDNDKSPSFIQIRSAKHKPTWVHYIAAEEEDWDYAPLVLA 420
 QY 421 PDRSYKSQYLNNGPQIRGKYKKVRPMAYTDETFKTREAIQHESGILGPLLYGEVDTL 480
 DB 421 PDRSYKSQYLNNGPQIRGKYKKVRPMAYTDETFKTREAIQHESGILGPLLYGEVDTL 480
 QY 481 LIIFKNOASRPYNIYPHGITDVRLPYSRRLPKGVKHLKDFPILPGEIFKFKWTVTVDGP 540
 DB 481 LIIFKNOASRPYNIYPHGITDVRLPYSRRLPKGVKHLKDFPILPGEIFKFKWTVTVDGP 540
 QY 541 TKSDPRCLTRYSSFVNMRDLASGLIGPLLYCYKESVDQGNQIMSDKXNVILFSVFDE 600
 DB 541 TKSDPRCLTRYSSFVNMRDLASGLIGPLLYCYKESVDQGNQIMSDKXNVILFSVFDE 600
 QY 601 NRSWYLTENTQRELPNPAGVQLEDPEFQASNIHSGYVFDLSQLSVCLHEVAYWYLS 660
 DB 601 NRSWYLTENTQRELPNPAGVQLEDPEFQASNIHSGYVFDLSQLSVCLHEVAYWYLS 660
 QY 661 IGAQTDLSVFFSGYTFKHMYEDTTLTPFSGETVPMGNEPGLMILGCHNSDFNRG 720
 DB 661 IGAQTDLSVFFSGYTFKHMYEDTTLTPFSGETVPMGNEPGLMILGCHNSDFNRG 720
 QY 721 MTALLKVSSCKNTGDIYEDYSIEDISAYLLSKNNAIEPRSFQNSRHPSTROKQFNATTI 780
 DB 721 MTALLKVSSCKNTGDIYEDYSIEDISAYLLSKNNAIEPRSFQNSRHPSTROKQFNATTI 780
 QY 781 PENDIEKTPWFAHRTMPKIQNVSSDLMRLQSPTPHGLSLSDLOEAKYETFSDDPS 840
 DB 781 PENDIEKTPWFAHRTMPKIQNVSSDLMRLQSPTPHGLSLSDLOEAKYETFSDDPS 840
 QY 841 PGALDSNNLSSEMTHERPQLHSGDMVTFPSGQLRLNEKLGTTAATLKKLDFKVSST 900
 DB 841 PGALDSNNLSSEMTHERPQLHSGDMVTFPSGQLRLNEKLGTTAATLKKLDFKVSST 900
 QY 901 SNNLIISTPSDNLAAAGTNTSSLGPPSMVHYDSQDITTLFGKKSPLTESGGPLSLEE 960
 DB 901 SNNLIISTPSDNLAAAGTNTSSLGPPSMVHYDSQDITTLFGKKSPLTESGGPLSLEE 960
 QY 961 NNDSKLLESGLMNSQESWGKNVSTESGLFKGRAGPALLTKDNALFKVSI SLKTN 1020
 DB 961 NNDSKLLESGLMNSQESWGKNVSTESGLFKGRAGPALLTKDNALFKVSI SLKTN 1020
 QY 1021 KTSNNSATNRKTHIDGPSLLIENSFVWQNILEDSTBFKVTPLIHDRMLMDKNATALRL 1080
 DB 1021 KTSNNSATNRKTHIDGPSLLIENSFVWQNILEDSTBFKVTPLIHDRMLMDKNATALRL 1080
 QY 1081 NHMKNKTTSSKNMNVQCKEGPIPPDAQNPDMSFFKMLFLPESARWIQTHGKNSLNSG 1140
 DB 1081 NHMKNKTTSSKNMNVQCKEGPIPPDAQNPDMSFFKMLFLPESARWIQTHGKNSLNSG 1140
 QY 1141 QGSPKQVLSLGPESKVEGQNFLEKXKVVVGKEFTKDVGLKEMWPPSSNLFITLNDN 1200
 DB 1141 QGSPKQVLSLGPESKVEGQNFLEKXKVVVGKEFTKDVGLKEMWPPSSNLFITLNDN 1200
 QY 1201 LHENNTNQEKKIOEEIEKKETLQIENWVLPQIHVATGTKNFMKNLFLSTRQNVESYD 1260
 DB 1201 LHENNTNQEKKIOEEIEKKETLQIENWVLPQIHVATGTKNFMKNLFLSTRQNVESYD 1260
 QY 1261 GAYAPVLDQFRLNDSTNRKTKHTAHFKSGEENLEGLGNQTKQIVEKACTRI SPNT 1320
 DB 1261 GAYAPVLDQFRLNDSTNRKTKHTAHFKSGEENLEGLGNQTKQIVEKACTRI SPNT 1320
 QY 1321 SQQFVTVQSRKALKQFRLPLEETELKRIIVDDTSTQWSKNMKHLTPSTETQIDYNEKE 1380
 DB 1321 SQQFVTVQSRKALKQFRLPLEETELKRIIVDDTSTQWSKNMKHLTPSTETQIDYNEKE 1380

```
QY 1381 KGAIQSPSLDCLTRSHSIPOANESPLIAKVSFBSIRPIYLTRVLFDQNSSHLPAASY 1440
DB 1381 KGAIQSPSLDCLTRSHSIPOANESPLIAKVSFBSIRPIYLTRVLFDQNSSHLPAASY 1440
QY 1441 RKDQSGVQESSHFLQGAQKNNLSLAILTLEMTGQREVSGISGTSATNSVTKKVENTVLP 1500
DB 1441 RKDQSGVQESSHFLQGAQKNNLSLAILTLEMTGQREVSGISGTSATNSVTKKVENTVLP 1500
QY 1501 KPDLPKTSKVELLPKVHIYQKDLFPPTETNSGSGHLDLVEGSLLOGTEGAIKWNEANRP 1560
DB 1501 KPDLPKTSKVELLPKVHIYQKDLFPPTETNSGSGHLDLVEGSLLOGTEGAIKWNEANRP 1560
QY 1561 GKVPFLRVATSSAKPSKLLDPLAWDNHYGTQPKSEWKSOEKSPEKTAFAKKDITLSL 1620
DB 1561 GKVPFLRVATSSAKPSKLLDPLAWDNHYGTQPKSEWKSOEKSPEKTAFAKKDITLSL 1620
QY 1621 NACSNHAIAINEGONKPEIEVTWAKQGRTERLCSQNPVVKRQHRBITRITLOSQDEE 1680
DB 1621 NACSNHAIAINEGONKPEIEVTWAKQGRTERLCSQNPVVKRQHRBITRITLOSQDEE 1680
QY 1681 IDYDTSIVEMKKEDFDIYDENQSPRSFQKTRHYTIAAVERLWDYCMSSSPHVLNR 1740
DB 1681 IDYDTSIVEMKKEDFDIYDENQSPRSFQKTRHYTIAAVERLWDYCMSSSPHVLNR 1740
QY 1741 AQSGSVPOFKKVVQFQFDDGSGTQPLXGELNEHGLLGPVIRAEVEDNIMVTFNQASR 1800
DB 1741 AQSGSVPOFKKVVQFQFDDGSGTQPLXGELNEHGLLGPVIRAEVEDNIMVTFNQASR 1800
QY 1801 PYSFYSSLISYEEDQQAEPKPNFVKNETITYFKVQHHVAPTKDFDCKAWAYFSDV 1860
DB 1801 PYSFYSSLISYEEDQQAEPKPNFVKNETITYFKVQHHVAPTKDFDCKAWAYFSDV 1860
QY 1861 DLEKDVHSLGCLPLVCHTNTLNPAHQGVTVQBEALPFTIIDEKTSKVFTEYNERNCA 1920
DB 1861 DLEKDVHSLGCLPLVCHTNTLNPAHQGVTVQBEALPFTIIDEKTSKVFTEYNERNCA 1920
QY 1921 PCNIQMEPTFKENYRFAHNGIYIMDTPLGLVMAQDQIRWYLLSGNSNENIHSIFSGH 1980
DB 1921 PCNIQMEPTFKENYRFAHNGIYIMDTPLGLVMAQDQIRWYLLSGNSNENIHSIFSGH 1980
QY 1981 VFTVRKKEBKVALNLYPGVETVEMLPKSGAGIWRVECLIGELHLAGMSTLFLVYSNKC 2040
DB 1981 VFTVRKKEBKVALNLYPGVETVEMLPKSGAGIWRVECLIGELHLAGMSTLFLVYSNKC 2040
QY 2041 QPLGWSGHIRDFQITASQGVQWAPKLARLHYSGSINAWSTKPPSWIKVDLLAPMII 2100
DB 2041 QPLGWSGHIRDFQITASQGVQWAPKLARLHYSGSINAWSTKPPSWIKVDLLAPMII 2100
QY 2101 HGIKTQGARQKPSLSLYISOFIIMYSLDGKKWQTYRGNSTGTLMVFFGNVDSSGIKHNIEN 2160
DB 2101 HGIKTQGARQKPSLSLYISOFIIMYSLDGKKWQTYRGNSTGTLMVFFGNVDSSGIKHNIEN 2160
QY 2161 PPIIARYILHPTHYSIRSTLAMELNGCDLNSCMPLGMESKAISDAQITASSYFTNMEA 2220
DB 2161 PPIIARYILHPTHYSIRSTLAMELNGCDLNSCMPLGMESKAISDAQITASSYFTNMEA 2220
QY 2221 TWSPSKARLHLOGRSNARWQVNNPKXEWLQVDFQKTMKVTVGTTQGVKSLTSMYVKEFL 2280
DB 2221 TWSPSKARLHLOGRSNARWQVNNPKXEWLQVDFQKTMKVTVGTTQGVKSLTSMYVKEFL 2280
QY 2281 ISSSQDGHQWTLFFQNGKVKVFGNQDSFTPVVNSLDPPLLTLYLRIHQSWVHQIALEM 2340
DB 2281 ISSSQDGHQWTLFFQNGKVKVFGNQDSFTPVVNSLDPPLLTLYLRIHQSWVHQIALEM 2340
QY 2341 EVLGCEAQDILY 2351
DB 2341 EVLGCEAQDILY 2351
```

RESULT 14

AAW11461

ID AAW11461 standard; protein; 2351 AA.

XX

```
AC AAW11461;
XX
DT 20-NOV-1997 (first entry)
XX
DE Active Factor VIII:C analogue V1717X.
XX
KW Factor VIII:C; analogue; glycoprotein; blood coagulation cascade;
KW fibrinogen; fibrin clot; haemostasis; haemophilia A; bleeding diathesis;
KW plasma protease; thrombin; immunogen; antibody; haemophiliac; therapy;
KW proteolytic cleavage.
XX
OS Homo sapiens.
OS Synthetic.
XX
FH Key
FT Peptide 1..19 Location/Qualifiers
FT Protein /note= "signal peptide"
FT Region /note= "mature Factor VIII:C"
FT Domain /note= "heavy chain fragment"
FT Region /note= "B domain"
FT Region /note= "light chain fragment"
FT Modified-site 1736
XX
PN WO9703195-A1.
XX
PD 30-JAN-1997.
XX
PF 09-JUL-1996; 95WO-US011444.
XX
PR 11-JUL-1995; 95US-0001025P.
XX
PS (CHIR ) CHIRON CORP.
XX
PI Hung DT, Cohen PE, Innis M;
XX
DR WPI; 1997-119050/11.
XX
PT Factor VIII:C analog modified adjacent to a non-activating Arg residue -
PT used in the treatment of haemophiliacs, by improvement of haemostasis.
XX
PS Claim 36; page; 90pp; English.
XX
CC AAW11330-W11472 represent active Factor VIII:C analogues of the
CC invention. These sequences were created by mutating the wild type Factor
CC VIII:C coding sequence (see AAW1357) using mutagenic primers. The
CC analogues comprise a native Factor VIII:C polypeptide modified at a site
CC adjacent to a non-activating Arg residue so that a Arg-Pro or Pro-Arg
CC dipeptide is created. Factor VIII:C is a large glycoprotein that
CC participates in the blood coagulation cascade that ultimately converts
CC soluble fibrinogen to insoluble fibrin clot, effecting haemostasis. A
CC deficiency in Factor VIII:C is responsible for haemophilia A, which is an
CC X-chromosome-linked inherited bleeding diathesis. Factor VIII:C is
CC activated by plasma proteases, such as thrombin. During activation the
CC mature polypeptide is cleaved to generate heavy and light chain fragments
CC that are further cleaved. Complexes of two or more of the analogues,
CC nucleic acids and vectors encoding them may be used alone or in
CC conjunction with each other, for the prevention or treatment of active
CC Factor VIII:C deficiency in a mammal. The analogues may be used as
CC immunogens to raise antibodies, and in the treatment of haemophiliacs, by
CC improvement of haemostasis. The analogues are resistant to proteolytic
CC cleavage and display increased plasma half-life. They may be administered
CC at lower dosages and by different modes of administration
XX
SQ Sequence 2351 AA;
```

Query Match

Best Local Similarity 99.9%; Score 12405; DB 2; Length 2351;

Matches 2349; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 MQLELSTCFCLLRFCFSAIRYIYLGAVELSDWYMQSDILGELPVDARFPRRYPKSPFN 60
Db 1 MQLELSTCFCLLRFCFSAIRYIYLGAVELSDWYMQSDILGELPVDARFPRRYPKSPFN 60
QY 61 TSUVYKKTLEFVEFTDHLFNTAKRPPNMWGLLPPTIOAEVYDVTVTILKNWASHPVSLEAV 120
Db 61 TSUVYKKTLEFVEFTDHLFNTAKRPPNMWGLLPPTIOAEVYDVTVTILKNWASHPVSLEAV 120
QY 121 GVSWKASEGAEDDQTSOREKEDDKVFPQGSHTYVWQVTKENGSPASDPLCLTYSYLSH 180
Db 121 GVSWKASEGAEDDQTSOREKEDDKVFPQGSHTYVWQVTKENGSPASDPLCLTYSYLSH 180
QY 181 VDLVKDINSGLIGALLVCRGSLAKETOTLHKELLIFAVDEGKSWHSTKNSLMODRD 240
Db 181 VDLVKDINSGLIGALLVCRGSLAKETOTLHKELLIFAVDEGKSWHSTKNSLMODRD 240
QY 241 AASARAWPKVHTNVCYNNRSLPGLIGCHRKSVMYMEVIMGTTPEVHSIFLEGHTFLVRNH 300
Db 241 AASARAWPKVHTNVCYNNRSLPGLIGCHRKSVMYMEVIMGTTPEVHSIFLEGHTFLVRNH 300
QY 301 ROASLEISPTFLTAQLLMDLGOFLSLSHCHDGMAYVYKVDSCPEBPQRMKONE 360
Db 301 ROASLEISPTFLTAQLLMDLGOFLSLSHCHDGMAYVYKVDSCPEBPQRMKONE 360
QY 361 EAEYDDDLTDSEMDVRFDDNSPFTQCBVAKGPKTWVHIAAREEDWDYAPLVLA 420
Db 361 EAEYDDDLTDSEMDVRFDDNSPFTQCBVAKGPKTWVHIAAREEDWDYAPLVLA 420
QY 421 PDDRSYKSOYLNNGPORIGRYKVRFMAYDETFTKREAIQHESGILGPLLYGEVGDTL 480
Db 421 PDDRSYKSOYLNNGPORIGRYKVRFMAYDETFTKREAIQHESGILGPLLYGEVGDTL 480
QY 481 LIIFKQOASRYNYPHGITDVRPLYSRRLPKGVKHLKDPILPGBIPKYKWTVVEDGP 540
Db 481 LIIFKQOASRYNYPHGITDVRPLYSRRLPKGVKHLKDPILPGBIPKYKWTVVEDGP 540
QY 541 TKSDPRCLTRYSSPVNMRDIASLIGLGLIICYKESVDQGNQVMSKRNVLIFSVEDE 600
Db 541 TKSDPRCLTRYSSPVNMRDIASLIGLGLIICYKESVDQGNQVMSKRNVLIFSVEDE 600
QY 601 NRSWYLTENIORFLPNPAGVLEDEPFOASNTMHSINGYVFDLSQLSVCLHEVAYWYILS 660
Db 601 NRSWYLTENIORFLPNPAGVLEDEPFOASNTMHSINGYVFDLSQLSVCLHEVAYWYILS 660
QY 661 IGAQTDFTSVFPGSYTFKHKWYEDTLTFPPSGETVPMKSNBPNGLNWLCHNSDFNRG 720
Db 661 IGAQTDFTSVFPGSYTFKHKWYEDTLTFPPSGETVPMKSNBPNGLNWLCHNSDFNRG 720
QY 721 MTALLKVSCKNTGQYEDSYEDISAYLLSKNNAIEPRSPQNGRHPSTROKQFNATTI 780
Db 721 MTALLKVSCKNTGQYEDSYEDISAYLLSKNNAIEPRSPQNGRHPSTROKQFNATTI 780
QY 781 PENDIEKTPWFAHRTMPKIKQNVSSDILLMLLRQSPTHGLSLSDLOEAKYETESDDPS 840
Db 781 PENDIEKTPWFAHRTMPKIKQNVSSDILLMLLRQSPTHGLSLSDLOEAKYETESDDPS 840
QY 841 PGAIIDNNLSLNTWFRPOLHSGDMWFTPEGLQIRLNEKLGTTAAATELKKLDFKVSST 900
Db 841 PGAIIDNNLSLNTWFRPOLHSGDMWFTPEGLQIRLNEKLGTTAAATELKKLDFKVSST 900
QY 901 SNNLISTIPSDNLAAGTDNTSISGPPSPVHVDSOLDTTLFGKKSPLTESGGPLSLGRE 960
Db 901 SNNLISTIPSDNLAAGTDNTSISGPPSPVHVDSOLDTTLFGKKSPLTESGGPLSLGRE 960
QY 961 NNDKSLLESGLMNSQESSMGKNVSTESGRLPKGKAHPALLTKDNALFKVSIISLLKTN 1020
Db 961 NNDKSLLESGLMNSQESSMGKNVSTESGRLPKGKAHPALLTKDNALFKVSIISLLKTN 1020
QY 1021 KTSNNSATNRKTHIDGFSLLIENSPPWQNIILSDTEPKKVTPLIHDRMLMDKNATALL 1080
Db 1021 KTSNNSATNRKTHIDGFSLLIENSPPWQNIILSDTEPKKVTPLIHDRMLMDKNATALL 1080

QY 1081 NEMSNTTSSKNMEMVQOKKEGPIPDQAQNDKSMFFKMLFLPESARWIOETHGKNLSNG 1140
Db 1081 NEMSNTTSSKNMEMVQOKKEGPIPDQAQNDKSMFFKMLFLPESARWIOETHGKNLSNG 1140
QY 1141 QGSPKQLVSLGPEKSVGQNFILSEKXNVVVGKEFTKOVGLKEMVPPSRNLPLTNLDN 1200
Db 1141 QGSPKQLVSLGPEKSVGQNFILSEKXNVVVGKEFTKOVGLKEMVPPSRNLPLTNLDN 1200
QY 1201 LHENNTNQEKKIOBETTEKKTILIOENVLPOJHTVTGTKNEMKULFLLSTRQNVESYD 1260
Db 1201 LHENNTNQEKKIOBETTEKKTILIOENVLPOJHTVTGTKNEMKULFLLSTRQNVESYD 1260
QY 1261 GAYAPVLQDPRSLNDSNTRTKHTAHFSKKGEBENLEGLGNQTKQIIVEKYACTRISPT 1320
Db 1261 GAYAPVLQDPRSLNDSNTRTKHTAHFSKKGEBENLEGLGNQTKQIIVEKYACTRISPT 1320
QY 1321 SQQNFVQSRKRALKQPRLPLESTELBKRIIVDDTSTQWKNMKHLTPSLTQIDYNEKE 1380
Db 1321 SQQNFVQSRKRALKQPRLPLESTELBKRIIVDDTSTQWKNMKHLTPSLTQIDYNEKE 1380
QY 1381 KGAITQSPDLCDLTRSHSIPOANESPLPIAKVSFPPIRPIYILTRVLFQDNSSHLPAASY 1440
Db 1381 KGAITQSPDLCDLTRSHSIPOANESPLPIAKVSFPPIRPIYILTRVLFQDNSSHLPAASY 1440
QY 1441 RKXDSGVQESSHFLQGAKKNNLSLAILTLEMTGDQREVGLGTSATNSVTYKKVENTVLP 1500
Db 1441 RKXDSGVQESSHFLQGAKKNNLSLAILTLEMTGDQREVGLGTSATNSVTYKKVENTVLP 1500
QY 1501 KPDLPKTSKGVKELLPKVHIYQKDLFPPTETNSGSPGHLDLVEGSLLOGTEGAIKWNEANRP 1560
Db 1501 KPDLPKTSKGVKELLPKVHIYQKDLFPPTETNSGSPGHLDLVEGSLLOGTEGAIKWNEANRP 1560
QY 1561 GKVPFLVATESSAKTPSKLLDPLANDNHVGTQIPKEEWSQESPKPTAPKKKDTILSL 1620
Db 1561 GKVPFLVATESSAKTPSKLLDPLANDNHVGTQIPKEEWSQESPKPTAPKKKDTILSL 1620
QY 1621 NACESNEHAIAINEGONKPEIEVTWAKQGRTERLCSONPVLKXHQREITRTTLQSDQEE 1680
Db 1621 NACESNEHAIAINEGONKPEIEVTWAKQGRTERLCSONPVLKXHQREITRTTLQSDQEE 1680
QY 1681 IDYDDTISVEMKKEDFDIYDEDENQSPRSPOKTRHYPIAAVERLWDYGMSSSHXUNR 1740
Db 1681 IDYDDTISVEMKKEDFDIYDEDENQSPRSPOKTRHYPIAAVERLWDYGMSSSHXUNR 1740
QY 1741 AOSGSVPQFKVVECEFTDGSFTQPLVGELENEHGLLGPVIRAEVEDNIWVTRNOASR 1800
Db 1741 AOSGSVPQFKVVECEFTDGSFTQPLVGELENEHGLLGPVIRAEVEDNIWVTRNOASR 1800
QY 1801 PYSVYSSLIISYEEDQORQGAEPKPNVKENETKYFWKVQHHMAPTKDBFDCWAYFSDV 1860
Db 1801 PYSVYSSLIISYEEDQORQGAEPKPNVKENETKYFWKVQHHMAPTKDBFDCWAYFSDV 1860
QY 1861 DLEKDVHSGLIGPLLVCHTNTLNBAHGRQVTVQSPALFTTFDETYSWVFENNERNCA 1920
Db 1861 DLEKDVHSGLIGPLLVCHTNTLNBAHGRQVTVQSPALFTTFDETYSWVFENNERNCA 1920
QY 1921 PCNIOMEDPTFKENYRPHAINGYIMDTLPGLVMAQDQIRWYLLSMGSGNENIHSIHSGH 1980
Db 1921 PCNIOMEDPTFKENYRPHAINGYIMDTLPGLVMAQDQIRWYLLSMGSGNENIHSIHSGH 1980
QY 1981 VFTVRKKSEYKXALYNLYPGVFETVEMLPKAGIWRVECLIGEHLHAGMSTFLVYSKNC 2040
Db 1981 VFTVRKKSEYKXALYNLYPGVFETVEMLPKAGIWRVECLIGEHLHAGMSTFLVYSKNC 2040
QY 2041 QTPGLWASGHIRDQITASGOYQWAPKLARLHVSIGSINAWSTKEPPSWIKVDLLAPMII 2100
Db 2041 QTPGLWASGHIRDQITASGOYQWAPKLARLHVSIGSINAWSTKEPPSWIKVDLLAPMII 2100
QY 2101 HGIKTQGARQFSSLYTSQFIIMYSLDGKKWOTYRGNSTGTIMWFFGVNDSGKHNIFN 2160
Db 2101 HGIKTQGARQFSSLYTSQFIIMYSLDGKKWOTYRGNSTGTIMWFFGVNDSGKHNIFN 2160
QY 2161 PPIIARYIRLPHYTHYSIRSTLNMELMGDLNSCMLPGWESKAIQITASSYFTNMPA 2220

Db 2161 PPIIARIIRLEPTYSIRKSTIRLMELMGCDLNSCMPJGSEKASDAQITASSFTNWA 2220
 QY 2221 TWSPSKARLHLQGRSNAWRPQNNPKREWLDVDFQKTMKVGTGTTQGVKSLLTSMYVKEFL 2280
 Db 2221 TWSPSKARLHLQGRSNAWRPQNNPKREWLDVDFQKTMKVGTGTTQGVKSLLTSMYVKEFL 2280
 QY 2281 ISSSODGHOWTLFFONGKVKVFOGNDSTFWKSLDPLLTRYLRIHPQSWVHQIALRM 2340
 Db 2281 ISSSODGHOWTLFFONGKVKVFOGNDSTFWKSLDPLLTRYLRIHPQSWVHQIALRM 2340
 QY 2341 EVLGEAODLY 2351
 Db 2341 EVLGEAODLY 2351

RESULT 15
 AAW11387
 ID AAW11387 standard; protein; 2351 AA.
 XX
 AC AAW11387;
 XX
 DT 18-NOV-1997 (first entry)
 XX
 DE Active Factor VIII:C analogue N357X.
 XX
 KW Factor VIII:C; analogue; glycoprotein; blood coagulation cascade;
 KW fibrinogen; fibrin clot; haemostasis; haemophilia A; bleeding diathesis;
 KW plasma protease; thrombin; immunogen; antibody; haemophiliac; therapy;
 KW proteolytic cleavage.
 XX
 OS Homo sapiens.
 OS Synthetic.
 XX
 FH Key Location/Qualifiers
 FT Peptide 1..19
 FT /note= "signal peptide"
 FT Protein 20..2351
 FT /note= "mature Factor VIII:C"
 FT Region 20..1667
 FT /note= "heavy chain fragment"
 FT Modified-site 376
 FT /label= Phe, Glu, Pro
 FT Domain 760..1667
 FT /note= "B domain"
 FT Region 1668..2350
 FT /note= "light chain fragment"
 XX
 PN WO9703195-A1.
 XX
 PD 30-JAN-1997.
 XX
 PF 09-JUL-1996; 96WC-US011444.
 XX
 PR 11-JUL-1995; 95US-0001025P.
 XX
 PA (CHIR) CHIRON CORP.
 XX
 PI Hung DT, Cohen FB, Innis M;
 XX
 XX MPI; 1997-119050/11.
 XX
 XX Factor VIII:C analog modified adjacent to a non-activating Arg residue -
 XX used in the treatment of haemophiliacs, by improvement of haemostasis.
 XX
 XX Claim 20; Page; 90pp; English.
 XX
 CC AAW11330-W11472 represent active Factor VIII:C analogues of the
 CC invention. These sequences were created by mutating the wild type Factor
 CC VIII:C coding sequence (see AAT51357) using mutagenic primers. The
 CC analogues comprise a native Factor VIII:C polypeptide modified at a site
 CC adjacent to a non-activating Arg residue so that a Arg-Pro or Pro-Arg
 CC dipeptide is created. Factor VIII:C is a large glycoprotein that

CC participates in the blood coagulation cascade that ultimately converts
 CC soluble fibrinogen to insoluble fibrin clot, effecting haemostasis. A
 CC deficiency in Factor VIII:C is responsible for haemophilia A, which is an
 CC X-chromosome-linked inherited bleeding diathesis. Factor VIII:C is
 CC activated by plasma proteases, such as thrombin. During activation the
 CC mature polypeptide is cleaved to generate heavy and light chain fragments
 CC that are further cleaved. Complexes of two or more of the analogues,
 CC nucleic acids and vectors encoding them may be used alone or in
 CC conjunction with each other, for the prevention or treatment of active
 CC Factor VIII:C deficiency in a mammal. The analogues may be used as
 CC immunogens to raise antibodies, and in the treatment of haemophiliacs, by
 CC improvement of haemostasis. The analogues are resistant to proteolytic
 CC cleavage and display increased plasma half-life. They may be administered
 CC at lower dosages and by different modes of administration
 XX
 SQ Sequence 2351 AA;
 Query Match 99.9%; Score 12405; DB 2; Length 2351;
 Best Local Similarity 99.9%; Pred. No. 0;
 Matches 2349; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 1 MQIELSTCFPF-CHLRFCFSATRYVLGAVELSWDYMQSDLGELPVDARFPPRVKSPFN 60
 Db 1 MQIELSTCFPF-CHLRFCFSATRYVLGAVELSWDYMQSDLGELPVDARFPPRVKSPFN 60
 QY 61 TSVVYKKTFLFVEFTDELFNIAKPRPPWMLLGPTIOAEVYDVTVTILKNMASHPVSLHAV 120
 Db 61 TSVVYKKTFLFVEFTDELFNIAKPRPPWMLLGPTIOAEVYDVTVTILKNMASHPVSLHAV 120
 QY 121 GVSVMKASEGAEYDDQTSQREKEDDKVPFGSHYVWQVLKENGFMASDPLCLTYSLSH 180
 Db 121 GVSVMKASEGAEYDDQTSQREKEDDKVPFGSHYVWQVLKENGFMASDPLCLTYSLSH 180
 QY 181 VDLVKDLSGLIGALLVCREGSLAKEKQTQHLKILLFAVDEGKSWHSEKNSLMQDRD 240
 Db 181 VDLVKDLSGLIGALLVCREGSLAKEKQTQHLKILLFAVDEGKSWHSEKNSLMQDRD 240
 QY 241 AASARAWPKMHTVNGYVNRSLPGLIGCHRKSVYWHVIGMGTTPPEVHSIFLGHTFLVRNH 300
 Db 241 AASARAWPKMHTVNGYVNRSLPGLIGCHRKSVYWHVIGMGTTPPEVHSIFLGHTFLVRNH 300
 QY 301 ROASLEISFIFTLTAQTLLMDLGOFLLSCHISSHOHDGMEAYVKVDSCEPEPQLEMKNE 360
 Db 301 ROASLEISFIFTLTAQTLLMDLGOFLLSCHISSHOHDGMEAYVKVDSCEPEPQLEMKNE 360
 QY 361 EADYDDDLTDSEMDVDFDDNSPFIQIRSVAKKHPTWVHYIAAEEEDWDYAPLVLA 420
 Db 361 EADYDDDLTDSEMDVDFDDNSPFIQIRSVAKKHPTWVHYIAAEEEDWDYAPLVLA 420
 QY 421 PDDRSYKSYLNNGPQRIGRYKVKVRFMAYTDETFKTRATQHSGLIGPLLYGVGDTL 480
 Db 421 PDDRSYKSYLNNGPQRIGRYKVKVRFMAYTDETFKTRATQHSGLIGPLLYGVGDTL 480
 QY 481 LLIIFKNQASRPNTYPHGITDVRPLYSRRLPKGVKHLKDFPLPGEIIFYKWTVTVEDGP 540
 Db 481 LLIIFKNQASRPNTYPHGITDVRPLYSRRLPKGVKHLKDFPLPGEIIFYKWTVTVEDGP 540
 QY 541 TKSDPRCLTRYSSFVNMRDLASGLIGPLLYCYKESVDQRGNQIMSDKRNVLFSVPE 600
 Db 541 TKSDPRCLTRYSSFVNMRDLASGLIGPLLYCYKESVDQRGNQIMSDKRNVLFSVPE 600
 QY 601 NRSWYLTEINIQBFLENPAGVQLEDEFOASNTMHSINGVYVFDLSQLSVCLHVAWYILS 660
 Db 601 NRSWYLTEINIQBFLENPAGVQLEDEFOASNTMHSINGVYVFDLSQLSVCLHVAWYILS 660
 QY 661 IGAQTDLSVFPFSGYTFKHMYVETLTLFPFSGTVMFMSMENPGLWILGCHNSFNRNG 720
 Db 661 IGAQTDLSVFPFSGYTFKHMYVETLTLFPFSGTVMFMSMENPGLWILGCHNSFNRNG 720
 QY 721 MTALLKVSSCDNTGYEDSDYEDISAYLLSKNNAIEPRFSQNRHSTROKQFNATTI 780
 Db 721 MTALLKVSSCDNTGYEDSDYEDISAYLLSKNNAIEPRFSQNRHSTROKQFNATTI 780

QY 781 PENDIEKTDWFAHRTWMPKIONVSSDLLMLLRQSPHPGLSLSDLQEAKEYETFSDDPS 840
DB 781 PENDIEKTDWFAHRTWMPKIONVSSDLLMLLRQSPHPGLSLSDLQEAKEYETFSDDPS 840
QY 841 PGATDSNNSISEMTHFRPOLHSGDMVFTPSGLOLRINEKLGTTAAATFELKLLDFKVSST 900
DB 841 PGATDSNNSISEMTHFRPOLHSGDMVFTPSGLOLRINEKLGTTAAATFELKLLDFKVSST 900
QY 901 SNNLISITIPSONLAAGTNTSSIGFPPSPVHYDQDITTLFGKKSSPLTSGGGLSLSBE 960
DB 901 SNNLISITIPSONLAAGTNTSSIGFPPSPVHYDQDITTLFGKKSSPLTSGGGLSLSBE 960
QY 961 NNDKSLLESGLMNSOESSGKNVNSTESGRFLFKGKRAHGPALLTKONALFKVSTSLAKTN 1020
DB 961 NNDKSLLESGLMNSOESSGKNVNSTESGRFLFKGKRAHGPALLTKONALFKVSTSLAKTN 1020
QY 1021 KTSNNSATNRKTHIDGPSLLIENSPPVQWONILESDTEPKVTPLIHRYLMDKXNATALRL 1080
DB 1021 KTSNNSATNRKTHIDGPSLLIENSPPVQWONILESDTEPKVTPLIHRYLMDKXNATALRL 1080
QY 1081 NHMKNKTTSSKNMENVQOKSGPIPPDAQNPMDFFKMLFLPESARWIQRTHGKNSLNSG 1140
DB 1081 NHMKNKTTSSKNMENVQOKSGPIPPDAQNPMDFFKMLFLPESARWIQRTHGKNSLNSG 1140
QY 1141 QGPPSPKQIVSLGPEKSVBQNFLEKKNVAVVGKEFTKDVGLKEMVFPSSRNLFITNLND 1200
DB 1141 QGPPSPKQIVSLGPEKSVBQNFLEKKNVAVVGKEFTKDVGLKEMVFPSSRNLFITNLND 1200
QY 1201 LHENNTHNQBKIOBEBIKKKTLLQENVVLQIHTVGTQKFMKMLFLLSTRQNVGGSYD 1260
DB 1201 LHENNTHNQBKIOBEBIKKKTLLQENVVLQIHTVGTQKFMKMLFLLSTRQNVGGSYD 1260
QY 1261 GAYAPVLQDPPSLNDSTNRKTHAHFSGKBEENLEGKNOTKQIVBKVACTTRISNT 1320
DB 1261 GAYAPVLQDPPSLNDSTNRKTHAHFSGKBEENLEGKNOTKQIVBKVACTTRISNT 1320
QY 1321 SQQNFVTOQRKALQKRLPLEETELEKRIIIVDDTSTQWSNMKHLTPSTLTQIDYNEKE 1380
DB 1321 SQQNFVTOQRKALQKRLPLEETELEKRIIIVDDTSTQWSNMKHLTPSTLTQIDYNEKE 1380
QY 1381 KGATQSPSLCLTRSHSIPQANRSPLPIAKVSSPSPSRPIYLTRVLFDQNSSHLPAASY 1440
DB 1381 KGATQSPSLCLTRSHSIPQANRSPLPIAKVSSPSPSRPIYLTRVLFDQNSSHLPAASY 1440
QY 1441 RKQSGVQOESSHETLGAKKNLSLAILILEMTGDQ3VGSIGTSATNSVYKVENTVLP 1500
DB 1441 RKQSGVQOESSHETLGAKKNLSLAILILEMTGDQ3VGSIGTSATNSVYKVENTVLP 1500
QY 1501 KPDLPKTSKGVVELLPKVHIYKDLPTTETSNGPSGHLDLVBGSLLOQTEGAIKWNEANRP 1560
DB 1501 KPDLPKTSKGVVELLPKVHIYKDLPTTETSNGPSGHLDLVBGSLLOQTEGAIKWNEANRP 1560
QY 1561 GKVPPLRVATESSAKTPSKLLDPLAWDNHYGTQIPKEBWSQEKSPKTAFAKKDITLSL 1620
DB 1561 GKVPPLRVATESSAKTPSKLLDPLAWDNHYGTQIPKEBWSQEKSPKTAFAKKDITLSL 1620
QY 1621 NACESNHAIAAINEGQNKPEIEVTWAKQRTBRLCSQNPVVLKRHQREBTRITTLQSDQEE 1680
DB 1621 NACESNHAIAAINEGQNKPEIEVTWAKQRTBRLCSQNPVVLKRHQREBTRITTLQSDQEE 1680
QY 1681 IDYDDTISVEMKKEDFDYDEDENSPSPSFQKTRHYFIAAVERLWDYGMSSSPHVLNR 1740
DB 1681 IDYDDTISVEMKKEDFDYDEDENSPSPSFQKTRHYFIAAVERLWDYGMSSSPHVLNR 1740
QY 1741 AOSGVSPOFKVVFQEFFTDGSGTQPLYRGELNEHGLGPIYRAEVEDNIMVTFRNQASR 1800
DB 1741 AOSGVSPOFKVVFQEFFTDGSGTQPLYRGELNEHGLGPIYRAEVEDNIMVTFRNQASR 1800
QY 1801 PYSFYSSLSISYEDORQGAEPKRFVKNETXTYFWKVOHMAPTKDEPDCWAYFSDV 1860
DB 1801 PYSFYSSLSISYEDORQGAEPKRFVKNETXTYFWKVOHMAPTKDEPDCWAYFSDV 1860
QY 1861 DLEKDVHESGLIGPLLVCHTNTLNPAHGRQVTVQEFALFTTIFDETKSWYFTENNERCRA 1920

DB 1921 DLEKDVHESGLIGPLLVCHTNTLNPAHGRQVTVQEFALFTTIFDETKSWYFTENNERCRA 1920
QY 1921 PCNIQMBDPTFKENYRPHAINGYIMDTPLGLVNAQDQIRIWAYLLSNGSNERNIHSIHFSGH 1980
DB 1921 PCNIQMBDPTFKENYRPHAINGYIMDTPLGLVNAQDQIRIWAYLLSNGSNERNIHSIHFSGH 1980
QY 1981 VFTVRKEEYKMALYNLYPGVFETVEMLPKSKAGIWRVECLIGEHLHAGMSTFLVYSNKC 2040
DB 1981 VFTVRKEEYKMALYNLYPGVFETVEMLPKSKAGIWRVECLIGEHLHAGMSTFLVYSNKC 2040
QY 2041 QTPLGMSAGHIRDFOITASGOYGOVAPKRLARLHYSGSINAWSTKEPFSWIKVDLLAPMII 2100
DB 2041 QTPLGMSAGHIRDFOITASGOYGOVAPKRLARLHYSGSINAWSTKEPFSWIKVDLLAPMII 2100
QY 2101 HGIKTQAGAROKFSSLYISOFTIMYSIDGKKWQTYGKSTGTLMVFFGNVDSGKHNIFN 2160
DB 2101 HGIKTQAGAROKFSSLYISOFTIMYSIDGKKWQTYGKSTGTLMVFFGNVDSGKHNIFN 2160
QY 2161 PPIIARVIRLHPHTHYSIRSTLRMELMGCDLNSCSMPLGMESKASDAQITASSYFTNMFA 2220
DB 2161 PPIIARVIRLHPHTHYSIRSTLRMELMGCDLNSCSMPLGMESKASDAQITASSYFTNMFA 2220
QY 2221 TWSPSKARLHLQGRSNARWPOVNNPKEWLQVDFOKTMMKVTVTTQGVKSLLTSMYKBEFL 2280
DB 2221 TWSPSKARLHLQGRSNARWPOVNNPKEWLQVDFOKTMMKVTVTTQGVKSLLTSMYKBEFL 2280
QY 2281 ISSODQGHQWTLFPONGKVYVFGQNDQSFPPVANSLOPPLLTRYLRIHPQSWVHOIALRM 2340
DB 2281 ISSODQGHQWTLFPONGKVYVFGQNDQSFPPVANSLOPPLLTRYLRIHPQSWVHOIALRM 2340
QY 2341 EVLGCEAQDLY 2351
DB 2341 EVLGCEAQDLY 2351

Search completed: April 13, 2004, 14:04:07
Job time : 77.5 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 13, 2004, 14:01:28 ; Search time 27 Seconds
(without alignments)
8375.788 Million cell updates/sec

Title: NP000123-328
Perfect score: 12416
Sequence: 1 MQEISLSTCFLLCLLRFCSA.....WVHQIALRMEVLGCEAQDLY 2351

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR 78: *
1: Piri: *
2: Pir2: *
3: Pir3: *
4: Pir4: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	12410	100.0	2351	1 EZHU	coagulation factor
2	8827	71.1	2319	2 A47004	coagulation factor
3	8068	65.0	2133	2 T42783	coagulation factor
4	2823.5	22.7	2224	1 KPEU5	coagulation factor
5	2774.5	22.3	2183	2 T42764	coagulation factor
6	2747	22.1	2211	1 KPEO5	coagulation factor
7	2190	17.6	859	2 A25945	coagulation factor
8	1326	10.7	1069	1 KTHU	ferroxidase (EC 1.11.1.1)
9	1276	10.3	1059	1 A35210	ferroxidase (EC 1.11.1.1)
10	1104	8.9	216	2 A44258	factor VIII-associ
11	663	5.3	427	2 J04915	ags protein precu
12	657	5.3	463	1 T436479	milk fat globule
13	650	5.2	409	2 T11743	pp47 protein - pig
14	635	5.1	401	2 S65138	glycoprotein anti
15	635	5.1	401	2 S74211	PAS-677 protein pr
16	443	3.6	927	1 J00948	A5 antigen precu
17	424.5	3.4	218	2 A47285	milk fat globule p
18	306.5	2.5	3133	2 S52093	hemocytin - silkw
19	262	2.1	845	2 JCS256	adipocyte transcri
20	239	1.9	1072	2 A86827	hypothetical prote
21	236.5	1.9	3418	1 G02334	breast cancer tumo
22	235.5	1.9	1358	2 A93360	SIR4 protein - yea
23	229	1.8	2166	2 G70163	hypothetical prote
24	224.5	1.8	3329	2 T30904	breast cancer tumo
25	223	1.8	2401	2 T28676	rhodopy protein -
26	223	1.8	2954	2 T14156	kinesin-related pr
27	222.5	1.8	3329	2 T42205	breast cancer susc
28	221	1.8	719	2 S51739	transcription repr
29	218.5	1.8	1928	2 S46773	myosin heavy chain

ALIGNMENTS

RESULT 1

EZHU
coagulation factor VIII precursor [validated] - human
N;Alternate names: antihemophilic factor A; coagulation factor VIIIc; procoagulant compo
C;Species: Homo sapiens (man)
C;Date: 28-Aug-1985 #sequence revision 28-Aug-1985 #text change 08-Dec-2000
C;Accession: I54318; A00525; I58059; A23584; A26174; A42348; A43986; S63527; S66445; B42
R;Gitschier, J.; Wood, W.I.
Hum. Mol. Genet. 1, 199-200, 1992
A;Title: Sequence of the exon-containing regions of the human factor VIII gene.
A;Reference number: I54318; MUID:93265012; PMID:1303178
A;Accession: I54318
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-1921, 'S', 1923-2351 <RES>
A;Cross-references: GB:M88648; NID:gi182381; PIDN:AAAS2420.1; PID:gi182383
R;Wood, W.I.; Capon, D.J.; Simonsen, C.C.; Eaton, D.L.; Gitschier, J.; Keyt, B.; Seaburg,
Nature 312, 330-337, 1984
A;Title: Expression of active human factor VIII from recombinant DNA clones.
A;Reference number: A00525; MUID:85061548; PMID:6438526
A;Accession: A00525
A;Molecule type: mRNA
A;Residues: 1-2351 <WOO>
R;Toole, J.J.; Knopf, J.L.; Wozney, J.M.; Sultzman, L.A.; Buecker, J.L.; Pittman, D.D.; I
S, D.N.; Hewick, R.M.
Nature 312, 342-347, 1984
A;Title: Molecular cloning of a cDNA encoding human antihemophilic factor.
A;Reference number: I58059; MUID:85061550; PMID:6438528
A;Accession: I58059
A;Status: nucleic acid sequence not shown; translated from GB/EMBL/DBJ
A;Molecule type: mRNA
A;Residues: 1-74, 'V', 76-1259, 'E', 1261-2351 <RE2>
A;Cross-references: GB:X01740; NID:gi182802; PIDN:AAAS2484.1; PID:gi182803
R;Truett, M.A.; Blacher, R.; Burke, R.L.; Caput, D.; Chu, C.; Dina, D.; Hartog, K.; Kuo,
B.; Randolph, A.; Urdea, M.S.; Valenzuela, P.; Dahl, H.H.; Favaliaro, J.; Hansen, J.; No;
DNA 4, 333-349, 1985
A;Title: Characterization of the polypeptide composition of human factor VIII:C and the
A;Reference number: A23584; MUID:86081164; PMID:3935400
A;Accession: A23584
A;Molecule type: mRNA
A;Residues: 1-2351 <TRU>
R;Eaton, D.; Rodriguez, H.; Vohar, G.A.
Biochemistry 25, 505-512, 1986
A;Title: Proteolytic processing of human factor VIII. Correlation of specific cleavages
ity.
A;Reference number: A26174; MUID:86159740; PMID:3082357
A;Accession: A26174
A;Molecule type: protein
A;Residues: 20-361392-399, 'X', 401-402; 1668-1678; 1709-1722, 'D', 1723-1725; 1741-1755 <EAT>
R;Pittman, D.D.; Wang, J.H.; Kaufman, R.J.

hypothetical prote
probable calcium-b
hypothetical prote
hypothetical prote
tpr homolog - fru
breast cancer tumo
amiloride-sensitiv
hypothetical coile
ankyrin 2, neurona
neurexin IV - fru
hypothetical prote
hypothetical prote
hypothetical prote
hypothetical prote
similar to late em
vacuolar protein V

Biochemistry 31, 3315-3325, 1992
 A>Title: Identification and functional importance of tyrosine sulfate residues within re
 A:Reference number: A42348; MUID:92207952; PMID:1554716
 A:Accession: A42348
 A:Molecule type: protein
 A:Residues: 20-36;356-372;392-408;582-594;1669-1699;1671;1672-1692;1693-1709;1709-17
 A:Experimental source: recombinant material from Chinese hamster ovary cells
 A>Note: sequence extracted from NCBI backbone and corrected to correspond with the publi
 R:Pay, P.J.; Smudzins, T.M.
 J. Biol. Chem. 264, 14005-14010, 1989
 A>Title: Intersubunit fluorescence energy transfer in human factor VIII.
 A:Reference number: A43986; MUID:89340500; PMID:2503509
 A:Accession: A43986
 A:Molecule type: protein
 A:Residues: 'X', 517-523;1853-1860, 'X', 1862-1864, 'X', 1866 <P>
 R:Leyte, A.; van Schijndel, H.B.; Niehrs, C.; Huttner, W.B.; Verbeet, M.P.; Mertens, K.;
 J. Biol. Chem. 266, 740-746, 1991
 A>Title: Sulfation of Tyr(1680) of human blood coagulation factor VIII is essential for
 A:Reference number: A56109; MUID:91033266; PMID:1898735
 A:Contents: annotation; sulfation
 R:Gitschier, J.; Wood, W.I.; Goralka, T.M.; Wion, K.L.; Chen, E.Y.; Eaton, D.H.; Vehar,
 Nature 312, 326-330, 1984
 A>Title: Characterization of the human factor VIII gene.
 A:Reference number: A56196; MUID:85061547; PMID:6438525
 A:Contents: annotation; introns
 R:McKullen, B.A.; Fujikawa, K.; Davie, E.W.; Hedner, U.; Ezban, M.
 Protein Sci. 4, 740-746, 1995
 A>Title: Locations of disulfide bonds and free cysteines in the heavy and light chains o
 A:Reference number: A56216; MUID:95338127; PMID:7613471
 A:Contents: annotation; disulfide bonds
 A>Note: 329-Cys, 711-Cys, and 2019-Cys were shown to have free sulfhydryls
 R:Kjalke, M.; Hedling, A.; Talbo, G.; Persson, E.; Thomsen, J.; Ezban, M.
 Eur. J. Biochem. 234, 773-779, 1995
 A>Title: Amino acid residues 721-729 are required for full factor VIII activity.
 A:Reference number: S63527; MUID:96163459; PMID:8575434
 A:Accession: S63527
 A:Molecule type: protein
 A:Residues: 733-752;753-759 <K>
 R:Lind, P.; Larsson, K.; Spira, J.; Sydow-Baeckman, M.; Almstedt, A.; Gray, E.; Sandberg
 Eur. J. Biochem. 232, 19-27, 1995
 A>Title: Novel forms of B-domain-deleted recombinant factor VIII molecules. Construction
 A:Reference number: S66445; MUID:96048024; PMID:7556150
 A:Accession: S66445
 A>Status: preliminary
 A:Molecule type: protein
 A:Residues: 1668-1685 <L>
 C:Comment: Factor VIII is activated by factor Xa and thrombin, but prolonged exposure pr
 C:Genetics:
 A:Gene: GDB:F8C
 A:Cross-references: GDB:119124; OMIM:306700
 A:Map position: Xq28-Xq28
 A:Introns: 48/2; 89/1; 130/1; 201/1; 224/1; 263/1; 337/1; 424/2; 481/3; 513/1; 584/3; 63
 A:Function:
 A>Description: acts as a cofactor, with calcium and phospholipid, for the factor IXa pro
 A:Pathway: Blood coagulation
 C:Superfamily: coagulation factor VIII; discoidin I amino-terminal homology; ferroxidase
 C:Keywords: acute phase; blood coagulation; duplication; glycoprotein; hemophilia A; pla
 F:1-19/Domain: signal sequence #status predicted <SIG>
 F:20-235/Domain: coagulation factor VIII #status experimental <MAT>
 F:20-740/Product: coagulation factor VIIIa heavy chain #status experimental <ACH>
 F:20-356/Domain: A1 <DA1>
 F:232-348/Domain: ferroxidase repeat homology <FOL>
 F:392-759/Domain: A2 <DA2>
 F:402-730/Domain: ferroxidase repeat homology <FOL>
 F:760-1667/Domain: B <DBO>
 F:1668-2351/Product: coagulation factor VIIIa light chain #status experimental <ACL>
 F:1709-2038/Domain: A3 <DA3>
 F:1716-2038/Domain: ferroxidase repeat homology <FOL>
 F:2039-2191/Domain: C1 <DC1>
 F:2039-2188/Domain: discoidin I amino-terminal homology <DNI>
 F:2192-2351/Domain: C2 <DC2>
 F:2192-2345/Domain: discoidin I amino-terminal homology <DNI>
 F:60-258, 601, 776, 803, 847, 919, 962, 982, 1020, 1024, 1074, 1085, 1204, 1274, 1278, 1301, 1319, 1403, 1

F:172-198, 267-348, 547-573, 649-730, 1851-1877, 1918-1922, 2040-2188/Disulfide bonds: #status
 F:355-356/Cleavage site: Arg-Met (coagulation factor Xa, protein C) #status predicted
 F:365, 737, 738, 742, 1693, 1699/Binding site: sulfate (Tyr) (covalent) #status experimental
 F:391-392/Cleavage site: Arg-Ser (coagulation factor Xa, thrombin) #status experimental
 F:414, 426/Binding site: sulfate (Tyr) (covalent) #status predicted
 F:753-760/Cleavage site: Arg-Ser (coagulation factor Xa, thrombin) #status experimental
 F:1667-1668/Cleavage site: Arg-Glu (unidentified proteinase) #status experimental
 F:1708-1709/Cleavage site: Arg-Ser (coagulation factor Xa, thrombin) #status experimental
 F:1740-1741/Cleavage site: Arg-Ala (coagulation factor Xa) #status experimental
 F:2193-2345/Disulfide bonds: #status predicted
 Query Match 100.0%; Score 12410; DB 1; Length 2351;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 2350; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 1 MQIELSCFFCLILRCPFSATRRYYLGAVELSWDMQSDLGELPVDARPPRRPKSPFPN 60
 DB 1 MQIELSCFFCLILRCPFSATRRYYLGAVELSWDMQSDLGELPVDARPPRRPKSPFPN 60
 QY 61 TSVVVKTLFVEFTDHLFNIAPKPPPMGLLPPTIOAEVDTVTITLKNMASHPVSLHAV 120
 DB 61 TSVVVKTLFVEFTDHLFNIAPKPPPMGLLPPTIOAEVDTVTITLKNMASHPVSLHAV 120
 QY 121 GVSVMKASEGAEYDDQTSOREKEDDKVFPFGSSHTYVMQVLKENGPMASDPLCLTYSYLSH 180
 DB 121 GVSVMKASEGAEYDDQTSOREKEDDKVFPFGSSHTYVMQVLKENGPMASDPLCLTYSYLSH 180
 QY 181 VDLVKDLSNGHIGALLVCREGLAKETQTLHKPILLFAVDEGKSHSETKNSLMQDRD 240
 DB 181 VDLVKDLSNGHIGALLVCREGLAKETQTLHKPILLFAVDEGKSHSETKNSLMQDRD 240
 QY 241 AASARAWPKMTVNGVYVNRSLPGILGCHKSVYWHVIGMTTPEVHSIFLEGHFPLVNRH 300
 DB 241 AASARAWPKMTVNGVYVNRSLPGILGCHKSVYWHVIGMTTPEVHSIFLEGHFPLVNRH 300
 QY 301 ROASLEISPIFLTAQTLMDLQGLFLSCHTSSHQHDMGMEAYVVDSCPEPQRLMKKNE 360
 DB 301 ROASLEISPIFLTAQTLMDLQGLFLSCHTSSHQHDMGMEAYVVDSCPEPQRLMKKNE 360
 QY 361 EABYDDDLTDEMDVVRFDNDSPFIQIRSVAKKPKTWVHYIAAEEEDWDVAPLVLA 420
 DB 361 EABYDDDLTDEMDVVRFDNDSPFIQIRSVAKKPKTWVHYIAAEEEDWDVAPLVLA 420
 QY 421 PDRSYKSOYLNNGPQRIGRKYKVRFWAYTDETFKTRAIQHSSEGLGLPYLGEVGDTL 480
 DB 421 PDRSYKSOYLNNGPQRIGRKYKVRFWAYTDETFKTRAIQHSSEGLGLPYLGEVGDTL 480
 QY 481 LIIFKNQASRPYNIYPHGIITVRPLYSRRLPKGVKHLKDPILPGEIPKYKWTVTVEGDP 540
 DB 481 LIIFKNQASRPYNIYPHGIITVRPLYSRRLPKGVKHLKDPILPGEIPKYKWTVTVEGDP 540
 QY 541 TKSDPRLCTRYSSFFVNMERDLASGLIGPLLI CYKESVDQGNQIMSKRVILFSVFDE 600
 DB 541 TKSDPRLCTRYSSFFVNMERDLASGLIGPLLI CYKESVDQGNQIMSKRVILFSVFDE 600
 QY 601 NRSWLTENIQRLFPNPAVQLEDPEPQASNMHSINGYVFDLSQLSVCLHEVAYWYLS 660
 DB 601 NRSWLTENIQRLFPNPAVQLEDPEPQASNMHSINGYVFDLSQLSVCLHEVAYWYLS 660
 QY 661 IGAQTDPLSVFSGYTFKHVMVYEDTLTLPFGSETVPMSENFGLWILGCHNDFRNRG 720
 DB 661 IGAQTDPLSVFSGYTFKHVMVYEDTLTLPFGSETVPMSENFGLWILGCHNDFRNRG 720
 QY 721 MTALLKVSSCKNTGDYEDSDYEDISAYLLSKNNAIEPRSSQNSRHPSTKQKFNATTI 780
 DB 721 MTALLKVSSCKNTGDYEDSDYEDISAYLLSKNNAIEPRSSQNSRHPSTKQKFNATTI 780
 QY 781 PENDIEKTDPPFAHRTPMKPIQNVSSESSDMLMLLRQSPFPHGLSLSDLOEAKYETFSDDPS 840
 DB 781 PENDIEKTDPPFAHRTPMKPIQNVSSESSDMLMLLRQSPFPHGLSLSDLOEAKYETFSDDPS 840
 QY 841 PCADSNNSLSEMTFRPOLHHSQDMVTFPSGQLRLNEKLGTTATTELKCLFKVSSST 900
 DB 841 PCADSNNSLSEMTFRPOLHHSQDMVTFPSGQLRLNEKLGTTATTELKCLFKVSSST 900

841 PGAIIDSNNSLSEMTHTPRPQLHGGDMVFTPESGQLRLNEKLTGTTAATLKLKDLFKVSST 900
901 SNHLI STIPSDNLAACTDNTSSIGPPSMPEHYDSOLDTTLFGKKSPLTSGGELSISEE 960
901 SNHLI STIPSDNLAACTDNTSSIGPPSMPEHYDSOLDTTLFGKKSPLTSGGELSISEE 960
961 NND SKLLESGLMNSQESSWCKNSVSTESGRLFKGKRAHGEFALLTKNALFKVSIILKTN 1020
961 NND SKLLESGLMNSQESSWCKNSVSTESGRLFKGKRAHGEFALLTKNALFKVSIILKTN 1020
1021 KTSNNSATNKTHTIDGSPLLIENSPPVQWNLSDTEFKKVTPLIHDRLMLDKNATAIRL 1080
1021 KTSNNSATNKTHTIDGSPLLIENSPPVQWNLSDTEFKKVTPLIHDRLMLDKNATAIRL 1080
1081 NMSNKTTSSKNMVMWCKEGEPTPPDAONPDSPFFKWLFLPESARWIQTHGKNSLNSG 1140
1081 NMSNKTTSSKNMVMWCKEGEPTPPDAONPDSPFFKWLFLPESARWIQTHGKNSLNSG 1140
1141 QGSPKQVSLGPEKSVGEGQNFSEKKNVVGKGEPTKDVGLKEMVPPSSRNJFLTNLDN 1200
1141 QGSPKQVSLGPEKSVGEGQNFSEKKNVVGKGEPTKDVGLKEMVPPSSRNJFLTNLDN 1200
1201 LHENNTNQBKKIQEIEKXETLIQENNVLPQIHVTGTGKXFMKQLPFLSTRQNVESYD 1260
1201 LHENNTNQBKKIQEIEKXETLIQENNVLPQIHVTGTGKXFMKQLPFLSTRQNVESYD 1260
1261 GAYAPVLQDFRSNDSTNRKKTATKSKGEEENLEGLNQTQKQIIVEKYACTTRISPNT 1320
1261 GAYAPVLQDFRSNDSTNRKKTATKSKGEEENLEGLNQTQKQIIVEKYACTTRISPNT 1320
1321 SQQNFVTRQSKRALQFRLPLEETLEKRIIVDDTSTQWKNMKHLTPSLTQIDYNEKE 1380
1321 SQQNFVTRQSKRALQFRLPLEETLEKRIIVDDTSTQWKNMKHLTPSLTQIDYNEKE 1380
1381 KGAIQTSPDLCDLTRSHSIPOANRSPPIAKVSPFPSPRIPIYTRVLPOQNSSHLPAASY 1440
1381 KGAIQTSPDLCDLTRSHSIPOANRSPPIAKVSPFPSPRIPIYTRVLPOQNSSHLPAASY 1440
1441 RKKDSGVQESHFPQGAKKNNLSAILEMTGQREVGSLGTATNSVTVKKVENTVLP 1500
1441 RKKDSGVQESHFPQGAKKNNLSAILEMTGQREVGSLGTATNSVTVKKVENTVLP 1500
1501 KPDLPKTSGKVELLPKHVHYQKOLFPPTETNSGSPGHLDLVEGSLLOQTEGAIKWNEANRP 1560
1501 KPDLPKTSGKVELLPKHVHYQKOLFPPTETNSGSPGHLDLVEGSLLOQTEGAIKWNEANRP 1560
1561 GKVPFLRVATESSAKTSPKLLDPLAWDNHYGTQPKGEWKSQEKSPKTAPEKKDITLSL 1620
1561 GKVPFLRVATESSAKTSPKLLDPLAWDNHYGTQPKGEWKSQEKSPKTAPEKKDITLSL 1620
1621 NACESNHAIAINRGONKPEIEVWAKQGRTERLCSONPVLKQHQREITRTTLQSDQEE 1680
1621 NACESNHAIAINRGONKPEIEVWAKQGRTERLCSONPVLKQHQREITRTTLQSDQEE 1680
1681 IDYDDTTISVMKKEDFDIYDEDENQSPRSFOKTRHYFIAAVERLWDYGMSSSPHVLNR 1740
1681 IDYDDTTISVMKKEDFDIYDEDENQSPRSFOKTRHYFIAAVERLWDYGMSSSPHVLNR 1740
1741 AQSGSVQFKKVVQEBTDSFTQPLRGELNEHLGLLGPVIRAEVDNIMWTRNQASR 1800
1741 AQSGSVQFKKVVQEBTDSFTQPLRGELNEHLGLLGPVIRAEVDNIMWTRNQASR 1800
1801 PYSFVSSLSIYEEORQABPKNFVKNPTKTVFWKQOHMAPTKDEFFDCKAWAYPSDV 1860
1801 PYSFVSSLSIYEEORQABPKNFVKNPTKTVFWKQOHMAPTKDEFFDCKAWAYPSDV 1860
1861 DLEKDVHSLGCLPLLVCHTNTLPAHGEQVTVQEFALPFTTIDETKSWYFTENMERNCR 1920
1861 DLEKDVHSLGCLPLLVCHTNTLPAHGEQVTVQEFALPFTTIDETKSWYFTENMERNCR 1920
1921 PCNTQMEDPTPKENYRFHAINGYINDTLPGLVMAQDQIRWYLLSYGNSNENIHHSFGH 1980
1921 PCNTQMEDPTPKENYRFHAINGYINDTLPGLVMAQDQIRWYLLSYGNSNENIHHSFGH 1980

1981 VFTVRKKEEYKALYNLYPGVPEVEMLPKAGIWEVECLIGELHAGMSTLFLVYSNKC 2040
1981 VFTVRKKEEYKALYNLYPGVPEVEMLPKAGIWEVECLIGELHAGMSTLFLVYSNKC 2040
2041 QTPGLMASGHIRDFOITASQGYQWAPKLARLHYSGSINAWSTKEPFSWIKVDLLAPMII 2100
2041 QTPGLMASGHIRDFOITASQGYQWAPKLARLHYSGSINAWSTKEPFSWIKVDLLAPMII 2100
2101 HGKTKGAKQKFSLSLISQFIIMYSLDGKKQTYRGNSGTGLMVFPGNVDSGIGKNIEN 2160
2101 HGKTKGAKQKFSLSLISQFIIMYSLDGKKQTYRGNSGTGLMVFPGNVDSGIGKNIEN 2160
2161 PPIIARIYRLHPHYSTIRSLRMLMGCCDINSCSMPLGMESKAISSAQITASSYFTNMFA 2220
2161 PPIIARIYRLHPHYSTIRSLRMLMGCCDINSCSMPLGMESKAISSAQITASSYFTNMFA 2220
2221 TWSPSKARLHLQGRSNAWRPQVNNPKWLQVDFQKMKVGTGVTQGVKSLTSMYVKEFL 2280
2221 TWSPSKARLHLQGRSNAWRPQVNNPKWLQVDFQKMKVGTGVTQGVKSLTSMYVKEFL 2280
2281 ISSQDGHQWTLFPQNGKVKVFGNQDSFTFPVNSLDPLLLTRYLRIHPQSWHQAIRLM 2340
2281 ISSQDGHQWTLFPQNGKVKVFGNQDSFTFPVNSLDPLLLTRYLRIHPQSWHQAIRLM 2340
2341 EVLGCBAQDLY 2351
2341 EVLGCBAQDLY 2351

RESULT 2
A47004
coagulation factor VIII precursor - mouse
C:Species: Mus musculus (house mouse)
C:Date: 09-Sep-1994 #sequence revision 09-Sep-1994 #text_change 18-Jun-1999
C:Accession: A47004
R:Elder, B.; Laskich, D.; Gitschier, J.
Genomics 16, 374-379, 1993
A:Title: Sequence of the murine factor VIII cDNA.
A:Reference number: A47004; MUID:93300511; PMID:8314577
A:Accession: A47004
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-2319 <BLD>
A:Cross-references: GR:U0573; NID:G192456; PID:AAA37385.1; PID:G192457
C:Superfamily: coagulation factor VIII; discoidin I amino-terminal homology; ferroxidase
F:1-19/Domain: signal sequence #status predicted <SIG>
F:23-349/Domain: ferroxidase repeat homology <FO1>
F:402-730/Domain: ferroxidase repeat homology <FO2>
F:1686-2006/Domain: ferroxidase repeat homology <FO3>
F:2007-2156/Domain: discoidin I amino-terminal homology <DN1>
F:2160-2313/Domain: discoidin I amino-terminal homology <DN2>

Query Match 71.1%; Score 8827; DB 2; Length 2319;
Best Local Similarity 72.5%; Pred. No. 0;
Matches 1712; Conservative 232; Mismatches 364; Indels 54; Gaps 22;

QY 1 MQELSTCFELCLLRFCEFSATRYIYGAVELSDWYQSD-LGELPYVDARFPRVPEKSPFP 59
DB 1 MQALPACFFLSFNFCSSAIRRIYGAVELSNWYIOSDLSLVHLDTRFLPRKSTSPFP 60
QY 60 NTSVYVYKTLTFVETDHLFNIAPRPPWMLGPTTQAEVYDVTITLKNWASHPVSJHA 119
DB 61 NTSIMYKTVFVEYKQOLFNIAPRPPWMLGPTTQAEVYDVTITLKNWASHPVSJHA 120
QY 120 VGVSYWKAASGAYDDQTSQREKDDKVPFGSGHTYVQVQLKENGPMASPLCTYSYLS 179
DB 121 VGVSYWKAASGAYDDQTSQREKDDKVPFGSGHTYVQVQLKENGPMASPLCTYSYMS 180
QY 180 HYDLVKDLSNGLGALLVCEGSLAKKQTLLHKEFTLLFAVDEGKSWHSETKNSLMQDR 239
DB 181 HYDLVKDLSNGLGALLVCEGSLAKKQTLLHKEFTLLFAVDEGKSWHSETKNSLMQDR 240

240 DAASARAWPMXHTVNGVYNRSLPGLIGCHRSKSVYVHVGTTPEVRSIFLECHTELVRN 299
241 DSASARAWPMXHTVNGVYNRSLPGLIGCHRSVYVHVGTTPEVRSIFLECHTELVRN 300
300 HQASLEISPIITFUAQTLLMDLQFLLSCHSSHQHDGVEAYVYKVDSCPEEPQLRMK-N 358
301 HQASLEISPIITFUAQTLLMDLQFLLSCHSSHQHDGVEAYVYKVDSCPEEPQKQNN 360
359 NEZEEDVDDLTDSMDVVRDDONSFPFOIRSVAKKHPTKWTWYTAZEEDWDYAPLV 418
361 NEEEDVDDDY-SEMDFTLDYDSSP-FIOIRSVAKKPYKWTWYTAZEEDWDYAPSV 418
419 LAPDRSYKSOYLNNGPORI GRKYKVKAFMAYTDETFKTRAIQHSIGLPLLYGEVGD 478
419 PTSNGSVKSOYLSNGPHRIGRKYKVKAFIAYTDETFKTRAIQHSIGLPLLYGEVGD 478
479 TLLIIFKQASRPYNIYPHGTTDVRPLYSRRLPGKVHKLKDPPLPGBEIFYKWTVTVED 538
479 TLLIIFKQASRPYNIYPHGTTDVRPLYSRRLPGKVHKLKDPPLPGBEIFYKWTVTVED 538
539 GPTKSDRCLTRYSSFFVNMERDLASGLIGPLLYCYKESVDORGNQIMSDKENVLESVF 598
539 GPTKSDRCLTRYSSFFVNMERDLASGLIGPLLYCYKESVDORGNQIMSDKENVLESVF 598
539 DENRSWYLTENIQRPVNPAGVQLEDZFOASNIMHSINGVDFSQVLSVCLHEVAYWYI 658
599 DENQSWYITENQRPVNPAGVQLEDZFOASNIMHSINGVDFSQVLSVCLHEVAYWYI 658
659 LSIQAOTPLSVFFSGYTFKHQWYEDTLTLPFSGETVFMSEMPGLWILGCNSDPRN 716
659 LSVGAOTPLSVFFSGYTFKHQWYEDTLTLPFSGETVFMSEMPGLWILGCNSDPRN 716
719 RGMPTALLKVSSCDXGTGYEDSYEDISAYLLSKNNAIEPRSFSONSRHPSPTROKFNAT 778
719 RGMPTALLKVSSCDXGTGYEDSYEDISAYLLSKNNAIEPRSFSONSRHPSPTROKFNAT 778
779 TIPENDIBKTDPMFAHRTMPKIQNVSSDMLMLLRQS-PPTHGLSLDLOQAKYETPDS 837
779 TIPKNDMEKIBEQEEIAEMLVQSVSVSDMLMLLQSGHPTPHGLFLSDGQEAIVEAIDH 838
838 DPSPCAIDSNLSLSEMTFRPOLHSGDGVFTPEGLQRLNEKLGTTAAETELKLDKFKV 897
839 DHSNPAIDSNEGPSKVTQLRPESHSEKXIVFTPQGLQSRNKSLETTIEVKWKKLGLQV 898
898 SSTSNMLI-STIPSNLAAGTNTSGLPSPMPVHYDSOLDTTLFGKSSSPSTESGGPLS 956
899 SSLPSNMLITILSDNLKATPEKTDSGPPDMPVHSSSKLSTTAGKAYSILVGHVPLN 959
957 LSEENNDKLESGLMNSQESGWKNVSGTESGRIPKGRAGPALLTKDNALFKVSTSL 1016
959 ASEENSDSNILDSITIMYSQESIPRNLIS IENDRLIREKRPFGIALLTKDNTLFRDNYSL 1018
1017 LKNTKTSNNSATNRKTHIDGSLILNENSPVWQV-ILESDETEKVTPLIHDRLMLDKNA 1075
1019 MKNTKTYHSTNEKHLTESPT-SENSTTDIQLAILKVNSEIQEVTALIHDTLLGKNS 1077
1076 TALRLNHNKNTTSKNMNMVQCKEGPTPPDAQNPDMSEFFKMLFLPSARMIQRTHGKN 1135
1078 TYVRLNHLNLRITSTXKNDI FHRKODDPIQDSENTIMPFSKMLFLSSSNWFKTNGN 1137
1136 SLNSGGPSKQVLSLGPESKVEGQNFLEKKNVVGKGEFTKVYGLKEMNFPSSRNPL 1195
1138 SLNSEOHSFPKQVILMFKKYVKNQSFLEKKNVVEQDGFKNIGLKOMAFPHNMSIFL 1197
1196 TNLNHLNHNTHNOEKKIOEBIEKKTLLIOENVVLPOHTVGTGNFMKNLFLSTRQNY 1255
1198 TNLNHLNHNTHNOEKKIOEBIE-KEALLEEKVLPQVHEA-GSKNFKLIDILIGTRQNI 1256
1256 EGSYDCAVAPVLQDFRSLNDSNRKTKKTAHFSEK-KGREENLEGIGNOTKQIVKYACT 1313
1257 --SLAYEVHVPVLQNITSINNSINTVQIHMEHFKEKRDKETNBSGLVKNKTREWVNY --- 1311
1314 TRISPNTSQCNFVTOQRKALQKFRJPLEETELEKRIIIVDDTSTQWSKNKXHLTSTLTQ 1373

RESULT 3

1312 -----PSQNIITQSRKALGOFRL-----STQWLKTNCSQTCIIKQ 1349
1374 IDYNEKKGAIITOSPISDCLTRSHSIPQANRSPPLIAKUSSPPSIRPIYITRVLFPQDNSS 1433
1350 IDHSKMKKEITKSSLSDS-SVIKSTTTQNSSDSHIVKTSAPP---PIDLKRFPFQNKFS 1405
1434 HLPAAVS---RKDSGVQESSFLOQAKKNLSLAILTLEMTGDQREVSGLSFSA NSV 1489
1406 HVQASSYIYDFKYSSRIQESNPLFKETKINNTSLALPNNMPIDQKFTSPGKSN NSV 1465
1490 TYKXVENTVLKPDLPKTSCKVELLPKVHIYQKDLFTETSNQSPGHLDLVEGSLLOQTE 1549
1466 TYKKRENIIFLKPTLPESGKIELLPQVSIQEBEILPTESSHGSPGHLNLMKVFLOKIQ 1525
1550 GAIKXNEANRPGVPFLRVATESAKTPSKLLDPLAWNHYGTQIPKEEWKSOQKSPEKT 1609
1526 GPTKWNKAKSHGB--S1KGXTESKNTRSKLNHAWDYHYAAQIPKDMWKSKEKSEI 1583
1610 AFKKKDTILS-NACESNHAIAINEGONKPEIETWAKQGRTERLCSQNPVPLKXHQREI 1669
1584 SIKGEDTILSRPHGNSHSIGA-NEKQWNPQRETTWVKQGTQRTCSQIPVPLKXHQREL 1642
1670 TRITLQSDQBEIDYDDTISVEMKKEDFDIYDEDNQSPRFQKTRHYFIAAVERIMDYG 1729
1643 --SAFQSEQATDYDDAITIE-TIEDEFIYSEDIKQSPRFQKTRHYFIAAVERIMDYG 1699
1730 MSSSPHVLNRAQSGSVPOPKKVVFOBTQSGFTQPLVYRGSNLNHLGLGPIYIRAEVEDN 1789
1700 MSTI-HVLRNRYQDNVFPQKVVVFQBTQSGFSQPLVYRGSNLNHLGLGPIYIRAEVEDN 1758
1790 IMVTFRQASRPYFYSLSIYSEDDQCGAEPKRNFKVKNETKTYFWKVQHIMAPTKDEF 1849
1759 IMVTFKQASRPYFYSLSIYSEDDQCGAEPKRNFKVKNETKTYFWKVQHIMAPTKDEF 1817
1850 DCKAWAFSDVDLEKDVHSGLIGPLLYCHNTNLPAGRQVTVQEPALFPPIFBETKSWY 1909
1818 DCKAWAFSDVDLEKDVHSGLIGPLLYCHNTNLPAGRQVTVQEPALFPPIFBETKSWY 1877
1910 FTEWERNCRAPCMIOEMEDPTEKENYRFAHNGYIMDTLPGLVMAQDQRIWYLLSMGNS 1969
1878 FTEWERNCRAPCMIOEMEDPTEKENYRFAHNGYIMDTLPGLVMAQDQRIWYLLSMGNS 1937
1970 ENIHSIFPSGHVFTVRKKEEYKMAVYLYPCVFTVEMLPSKAGIWRVVECLIGEHLHAGM 2029
1938 ENIHSIFPSGHVFTVRKKEEYKMAVYLYPCVFTVEMLPSKAGIWRVVECLIGEHLHAGM 1997
2030 STLFLVYSNKCOTPLGHASGHIRDPOITASQYQWAPKLAHLHYSGSINAWSTKEPESW 2089
1998 STLFLVYSNKCOTPLGHASGHIRDPOITASQYQWAPKLAHLHYSGSINAWSTKEPESW 2057
2090 IKVDLLADMIITHGKTQAGOKFSSLYISOFTIMVSLDGKKWQTVYRGNSTGTLVFFGCV 2149
2058 IKVDLLADMIITHGKTQAGOKFSSLYISOFTIMVSLDGKKWQTVYRGNSTGTLVFFGCV 2117
2150 DSSGKHNIFNPPIIARYIRLHPHTHSSIRSTRLMELMGCDLNSCMPJGMSKSAISDAQI 2209
2118 DSSGKHNIFNPPIIARYIRLHPHTHSSIRSTRLMELMGCDLNSCMPJGMSKSAISDAQI 2177
2210 TASSYFTNMPATWSPSKAHLQGSNAWRPOVNNPKENLOVDFQKTKMKTGVTTOGVKS 2269
2178 TASSYFTNMPATWSPSKAHLQGSNAWRPOVNNPKENLOVDFQKTKMKTGVTTOGVKS 2237
2270 LLTSYVKEFLISSSQDGHQWTLFFQNGKVKVFGQNDSTFPVWNSLDPPLLTLYRLRHP 2329
2238 LFTSMVKEFLISSSQDGHQWTLFFQNGKVKVFGQNDSTFPVWNSLDPPLLTLYRLRHP 2297
2330 QSWHQIALRMEVLGCEAQDLY 2351
2298 QIWEHQIALRMEVLGCEAQDLY 2319

T42763
coagulation factor VIII precursor - pig
C:Species: Sus scrofa domestica (domestic pig)
C:Date: 11-Jan-2000 #sequence_revision 11-Jan-2000 #text_change 09-Jun-2000
C:Accession: T42763
R:Lollar, P.
submitted to the EMBL Data Library, August 1996
A:Reference number: 22269
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-2133 <LOL>
A:Cross-references: EMBL:U49517; NID:G1511633; PID:G1511634; PIDN:AAB06703.1
C:Superfamily: coagulation factor VIII; disacoidin I amino-terminal homology; ferroxidase
C:Keywords: acute phase; blood coagulation; duplication; glycoprotein; hemophilia A; pla
F:1-19/Domain: signal sequence #status predicted <SIG>
F:20-2133/Product: coagulation factor VIII #status predicted <MA>
F:23-345/Domain: ferroxidase repeat homology <FOX1>
F:402-730/Domain: ferroxidase repeat homology <FOX2>
F:1498-1820/Domain: ferroxidase repeat homology <FOX3>

Query March 65.0%; Score 8068; DB 2; Length 2133;
Best Local Similarity 67.0%; Pred. No. 0;
Matches 157; Conservative 208; Mismatches 344; Indels 226; Gaps 17;

Qy 1 MQIELSTCFELCLLRFCSATRRYYLGAVELSWDMOSD-LGELPVDARPPRPVKPFP 59
Db 1 MQIELSTCFELCLLPLGFSAIRYYLGAVELSWYRQSELREHVDTRFPATPAGALPL 60

Qy 60 NTSVYKKTFLVEFTDLHFNIAKPRPMWGLGPTIQAEVYDTVITLKNWASHPVSIHA 119
Db 61 GPSVLKKTIVFVETDQLFVARPPPMWGLGPTIQAEVYDTVITLKNWASHPVSIHA 120

Qy 120 VGVSYWKASGAEVDDQTSQREKDDKVPFGSHTYVQVLKENGPMASDPLCTYSYLS 179
Db 121 VGVSWKASGAEVDDQTSQREKDDKVLPGKSTYVQVLKENGPMASDPLCTYSYLS 180

Qy 180 HVDLVKNLSGILGALLVCREGSLAKEKTQLHFKLILLFAVDEGKSWHSTKSLMQDR 239
Db 181 HVDLVKNLSGILGALLVCREGSLTRKTQNLHFEVLLFAVDEGKSWHNSANDSWTRAM 240

Qy 240 DAASARAWPKHVTNGYVNSLPLGLICHRKSVYWHVIGMTTPEVHSIFLEGHTFLVRN 299
Db 241 DPAPARAQPAHMTVNGYVNSLPLGLICHKSVYWHVIGMTTPEVHSIFLEGHTFLVRH 300

Qy 300 HQASLSISPTFTTAQTLMDLQGLLSCHISSHQDCMEAYVYKVDSCBPPEPOLRMKN 359
Db 301 HQASLSISPTFTTAQTLMDLQGLLLFCHISSHHGGMFAHVYVSCAEPQLRERAD 360

Qy 360 BEABDYDDDLTDSMDVVRFDNDSFPIQIRSVAKXHPKTVWHYIAAEEDMDYAPLVL 419
Db 361 BE-EDYDNDLYDSMDVVRFDGDDVSFPIQIRSVAKXHPKTVWHYISAEEDMDYAPAVP 419

Qy 420 APDRSKYQVLYNNGPQIRKYYKVPMAVTDFTFKTREAIQHSGLIGPLLYGVEGDT 479
Db 420 SPDSRSKSLVNSGPQIRKYYKARFVAVTDVTFKTRKAIPEYSGILGPLLYGVEGDT 479

Qy 480 LLIIIFKNQASRPYNIYPHGIITDVSLPYSRRLPKGVKHLKOPPLPGHIFKYYKWTVYEDG 539
Db 480 LLIIIFKNQASRPYNIYPHGIITDVSLPYSRRLPKGVKHLKOPPLPGHIFKYYKWTVYEDG 539

Qy 540 PTKSDPRLTRYSGSVFNMRDLASGITGPLLIYKVSVDQRGNQIMSDKRNVLFGVFD 599
Db 540 PTKSDPRLTRYSGSVINLEKDLASGLIGPLLIYKVSVDQRGNQIMSDKRNVLFGVFD 599

Qy 600 ENRSWYITENTQRLPNPAGVOLDEPQASNIHMSINGYVDSLQLSVCLHEVAYWYIL 659
Db 600 ENQSWYLAENTQRLPNPAGVOLDEPQASNIHMSINGYVDSLQLSVCLHEVAYWYIL 659

Qy 660 SIGACTPLSVFSCYTFKHMYVEDTLTLFPFSGETVPMSENPGLWILGCHNSDFRN 719
Db 660 SVGAQDTPLSVFSCYTFKHMYVEDTLTLFPFSGETVPMSENPGLWILGCHNSDFRN 719

720 GWTALLKVSODKNTGDIYEDSYEDI SAYLLSKNALEPRSFSONSRHPSTROKOFNATT 779
Db 720 GWTALLKVSODRIGDIYDNTYEDI PFGLLSGKNVLEPRSFQONSRPPSASQKQFQIT 779

780 IPENDIEKTPWFARHTPMPKIQNVSSDULMLLRQSPTHGLSLSDIQEAKYTFSDP 839
Db 780 SPEDDVE-LPQSGERTQALEELSVPDGMILLQNPAPHGSSSSDLOEARN--ADDY 836

780 SPAGIDNNLSLSEMTFRPOLHSGDMVFPESGLQLRLNEKLGTTAATELUKLDKFKVS 899
Db 837 LPGAERNTA?SAAARLRPELHSAERVLTP--EK-----ELKLDKMS 882

900 TSNNLIS--TIPSDNLAAGTNTSSIGPSPMPHYVDVDSOLDTTLFGKSSPJTBSGGPLSL 957
Db 883 SSDLLKTPITPSDLSAETRHSLGPPHPQVNFPSQLGAIVLGKSSSHFAGVPLGS 942

958 SEENDSKLLESGLMNSQSSWKNVSTESGRIFKGRAPGALLTXDNALFVSIIL 1017
Db 943 TEEDH-----ESSLGENVSPVESDGIKFEKRAHPASLTKDDVLFKVNISLV 989

1018 KTKTNNSATNRKTHIDGPSLLIENS?SVWQNLLESDTEPKKVTPLIHRMLMDKNATA 1077
Db 990 KTKARVYLKTRKIHIDDAALLTENRASA-----TFMDKNTTA 1028

1078 LRNLHMSNKTTSSKNVEMVQKKKGPIPPDAQNDPMDFKMLFLPESARLIQRTGKNSL 1137
Db 1029 SGLNHVN-----WIKGLGKNPL 1047

1138 NSGQSPKQVLSLGPESVGEQNFLEKNKVVGGEFTKDVGLKEMVFPSSNLELTN 1197
Db 1048 SSERGPSPELLTSSGSKSVKSGSGGGRIRVAVEEBELSGK--KEMMLPNSBLTFLN 1104

1198 LDMLEHNTINQEKLOEBIEKKETLLOENVLPOLHTVGTGNPMKQLFLLSTRQVVG 1257
Db 1105 SADVQGNDETSQKSEEMERREKLVQKVDLPQVYATGATGKTNFLNIHQSTEPSVEG 1164

1258 SYDGAVAPVLQDPRSLNDS?NRKHTAHFSKKEEENLEGLGNOTQIWEKYACTTRIS 1317
Db 1165 FDGSHAPVQDRLSNDLSAERASTHIAHSAIRERAPLAPGNRT----- 1210

1318 PNTSQQNFVTRQSKRALKQFRLPLEETEELKRIIVDDTSTQWSKNMKHLTPSTLTQIDYN 1377
Db 1211 -GPGPSRAVPRVKQSLQRLPLEEIKPERGVVNLATSRWS----- 1252

1378 EKSGAITQSPDLCDLTRSHSI?QANRSPPIAKVSPFPPIRYLTVLFDQNSHLPA 1437
Db 1253 ----- 1252

1438 ASYRKDSDGVQSSSHFIQGAKNLNLAILTLEMTGDQREVGSIGTSATNSVTVYKVENT 1497
Db 1253 -----ESSPILQGAKNLNLSPFLTLEWAGGQGGKISALGKSAAGPLASGKLEKA 1301

1498 VLPKPDLPKTSQKVELLPKHVHYOKOLFPPTETNGSPGHLDLVEGSLLOSTEGAUKNEA 1557
Db 1302 VLSAGLSEASGRAEFLPKVRVEREDLLPKTNSVNSCAHGLDGOEIEFLQTRGPVNLKV 1361

1558 NRCKVFPFLVAVETESSAKTSPKLLDPLAWNHYGTQIPKESWKSQESPEKTAFFKVDTI 1617
Db 1362 NRPG-----RTPSKLLGP-----PMPK-EWESLEKSPKSTALRTKDI 1398

1618 -LSLNACESNHAIAINEGQNKPEIEVTWAKQGRTERLCSQNPPLVLRHOREITRTTLOS 1676
Db 1399 SLPLDRHESNHSIAAKNEGGAET?CREAAWTQKQGPGLCAPKPPVLRHRHORDISLTFQP 1458

1677 DQBEIDVDDTIVEMKEDDFDIYDENQSPRSFQKTRHYFFIAAVERLWDYGWSSSPHV 1736
Db 1459 EEDKMDYDDIFSTETKGEDDFDIYGEDENQDPRSFQKTRHYFFIAAVERLWDYGWSSSPRA 1518

1737 LRNRQSGSVQPKKVVQFQFTPGSFTQPLRYGELNHLGLGPIYRAVEDNIMTFRN 1796
Db 1519 LRNRQNGEYPRFKVVFREFADGFTQPSYRGELNHLGLGPIYRAVEDNIMTFRN 1578

1797 QASRPYSFSLISYEDDSQGAEPKRN?FKVKNETKIYFWKVQHMAPTKDEDFCKAWAY 1856

Db 2053 TLRLLEQCEVNGCSTPLGMEKNGIKENKQITASSPFGKSWGDYWEPPARLNAAQGRVNAW 2112
 Qy 2239 RPQVNNKEMQLQVDFQKMTKVTGTTQGVKLLTSYMYKEFLISSQDCHQWTLFFQNGK 2298
 Db 2113 QAKANNKQWLEIDLLKFKKTAITQCKSLSEMYKVSITIHYSEQGVKWKPYRLKSS 2172
 Qy 2239 V--KVFOGQKDSFPVNVNSLDPPLTLRLIHPQSWVHQIALRMEVLGCEAQDLY 2351
 Db 2173 MVDKIFEGNTYTKGVKVKVFPFPIISRIVRIPK*WNSIALRLLELFC--DIY 2224
 RESULT 5
 T42764
 coagulation factor V - mouse
 C:Species: Mus musculus (house mouse)
 C>Date: 11-Jan-2000 #sequence_revision 11-Jan-2000 #text_change 09-Jun-2000
 C:Accession: T42764
 R:Yang, T.L.; Cui, J.; Rehmtulla, A.; Yang, A.; Moussalli, M.; Kaufman, R.J.; Ginsburg, Blood 91, 4593-4599, 1998
 A:Title: The structure and function of murine factor V and its inactivation by protein C
 A:Reference number: 222270; MUID:98282202; PMID:9616155
 A:Accession: T42764
 A>Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1-2183 <YAN>
 A:Cross-references: EMBL:U52925; NID:G3219690; PID:G3219691; PIDN:AAC99553.1
 C:Function:
 A:Pathway: blood coagulation
 C:Superfamily: coagulation factor V; discoidin I amino-terminal homology; ferroxidase re
 C:Keywords: blood coagulation; duplication; glycoprotein; phospholipid binding; plasma; P:350-682/Domain: ferroxidase repeat homology <FOX1>
 P:1541-1864/Domain: ferroxidase repeat homology <FOX2>
 Query Match 22.3%; Score 2174.5; DB 2; Length 2183;
 Best Local Similarity 29.9%; Pred. No. 8,1e-137; Indels 503; Gaps 73;
 Matches 750; Conservative 408; Mismatches 851;

Qy 526 EIFKYKVTWVEDQPTKSDPRCLTRYSSYFNMERDLASGLIGPLLIYKESVDQRNQI 585
 Db 477 ETFFYKNILFEDEPTENDAQCCLTPYSDVVDTRDIASGLIGLLICKRSLDQGVQR 536
 Qy 586 MSDKRNILSVFEDNSWYLTENIQFLNPAGVQLDEPFQASNMHISINGVYFSLQ 645
 Db 537 VADIEQAVFAVDENKSVIENKFCENPDVKGDDPKFYENSTMTINGVVPSSIS 596
 Qy 646 -LSVCLHEVAVWYILSJAQCTDFLSVPSGVTFKHKVYEDTLTLFSPSGTVMSENP 704
 Db 597 TLGFCFDDTVQWHEFCSVTHDDILTIHFTGHSFYIGRRHEDTLTLFPMRGESVTVTDNV 656
 Qy 705 GLWILGCHSNFRNRGMTALLKVSSCDKNTGDY-YEDSYE----- 743
 Db 657 GTWLTTWNSPKERNLRLRFDVKNR--DYDNEDSYBIYESPAPSTMTTRIHDSLE 713
 Qy 744 -----DISAVLLSKNNAIEPRFSQNSRRHSTKQKNATIP-END-EKTDWFA 793
 Db 714 NEFGIDNEDDDYQLLASSLG--RSPKNSLNP--EENEENLTALALENSEFISP-- 766
 Qy 794 HRTMPKIQNVSSDMLMLRQSPFPHGLSLSD-QEAKYETFSDDPSGA----- 843
 Db 767 ---STDRVVDNSRILSKIINN-----NLKQFRTL-----PGCAIVAGTLLNL 810
 Qy 844 --IDSNNLSLSEMTFRPOLHSGDMVFTPEGLQLRLNEKLGTTAATLKKLDKFKVSSTS 901
 Db 811 IGLDENFVLNSSTEHRSSSYHENDME-NPQSNITWVYLLPLGPKG----- 854
 Qy 902 NNLSTTPSDNLAAGTNTSSLGPPSPMPVHVDQI-DITL--FGKSSPLTBSGGPLSLSE 959
 Db 855 -----SGNREQDKPATIKTPHMKHFRFWMKAPAGKTGRHSPNPKNSYSG--MKSE 904
 Qy 960 ENDSKL-----LESGLMN-----SQESSW---GKNVSTESGRLFKGRAGPALLT 1004
 Db 905 EDISELPLQKZTSFELNRWRVASEKSYELIANGCEDTVDKLI----- 951
 Qy 1005 KDNALFKVSIILLKNTKNTSNNAIRKTHIDGPSLLIENSFVWQNLIESTEFKKVTPL 1064
 Db 952 -TNSPQNIITVPRGESHTNTTKPSDLPFTPSGVGHKSPhVQEQ--EENSQGKRLQF 1008
 Qy 1065 IHRMLDKNATALRLNHSNKITSSKNMENVQKKEGPIPPDAQNPDMSPFKMLFLPES 1124
 Db 1009 IRTI-----KKYKKNKAL-----HSPSPRGFDP----- 1033
 Qy 1125 ARWIQTHGKNSLNSGGQSPKQLVSLGPEXSVGQNFLSEKNKVVVGKEFTKDVGLKE 1184
 Db 1034 -----LEGHNHP----- 1041
 Qy 1185 MVFPSSRLFTJNLNDLHNNTHNQEKIQEIEKKETLIOENVVLPQI--HTVTGTKNP 1242
 Db 1042 --FPDRLLNHSLL--LHKSN-----ETALSPLANQTSPEMSTDRSLPDYNOYSKNDTSQM 1093
 Qy 1243 MKNFLPLSTRQNVGSDYGAVALVQDFRSLNDSNTNRKHTAHFSKKEEENL----- 1296
 Db 1094 SSSLDLY---QSVFAEHSPTFA-QDPDQTHSTIDPSYRSPPELSQGLDYDLSDHFP 1149
 Qy 1297 EGLGNQTKQIEVKYACTTRISPTNSQQNFVQSRKRALXQFRLPL--BETLEKRII--- 1351
 Db 1150 DDIG-----LTSFPDQSQKSSFSDDQQAIPSSDLSLFTISPELDQTIIPD 1197
 Qy 1352 VDDTSTQWKNKXHLTPSTLTQI-----DYNEK-----EKGAIQPSLDCLTRSHSIPQAN 1403
 Db 1198 LDQLLSPEDKQKTSSTD-LGQVFLSPDDNQKTSSTD/LGOVSLSPDDNQKTS---PDIG 1253
 Qy 1404 RSLPI---AKVSPSPSIRPIYLT-----RVLFQNSHLPAAVYKXKDSGVQESSHFLQ 1456
 Db 1254 QVFLSLDDNQKTS--PDIGQVFLSPDDNQKTSSTD/LGQVFLSPDDNQKTSSTD/LGQVFLSP 1312
 Qy 1457 AKKNLSLAIULTLEMTGQREVGSIGTSATNSVTYKVENTVLPKPD-----LPKTSKV 1511
 Db 1313 EDNQVFLDLQVPLSSDQNG---STSTDLTLSPDFGQTVLSPDLQQLPLSDNSQV 1368
 Qy 1512 ELLPKVH-----IYQKDLFPPTETNSGSPGHLDLVEGSLLOQTEGAIKWNEANRP 1560

```
Db 1369 TVSPDLSLLTSPDFNLIAPDLGQVTLSP-----DLQTN----- 1405
QY 1561 GKVPFLRVATSSAKTSPKLLDPLANDHNYGTQIPKEWKSQKSPKTKFAKXDTILSL 1620
Db 1406 ---PALNHGCHKASSADPDQASYDP--DSGQASSLP--ELAKTILPHPLDTHIPPSPSPPTL 1458
QY 1621 NACESNHAIAINEGQNKPIEVNTAKOGTERLCSQNPV---LKRHQREIRFTTILQSD 1677
Db 1459 N-----NTLSRKFNPLVWVGLSRVGDGVVEIPSE 1490
QY 1678 QBEIDYDDTISVEMKK3DFDIYDEDENQSPRSFOKTRH-----YFI 1719
Db 1491 PERLDED-----YAEDDFTYNDPYTDRTDYNSSRNPDTTAAWYLRGHGCHKFYI 1544
QY 1720 AAVERLDWYGMSSPHVLRNRAQSGVQP---FKVYVQFBFTGSGFTQPIYRGLNHLG 1776
Db 1545 AABBITWNAEFAQSEM--DHEDTGHTPKQTTVKVYKRYLDSTFTSRDPRAEYEBHLG 1602
QY 1777 LIGPYIRAEVDNLTWTFRNOASRPSPYSLSISYE-----EDQEQGAEPKFNKVP 1828
Db 1603 ILGPVIRAEVDVTVQVAFKMLASPPYSLHAHGLSYSEKSSGKTYEDBSPEWFODDAVQP 1662
QY 1829 NETKTYFWKQHHWAPKDFCDKAWAYFSDVLEKQVHSGLIGLPLVCHTINTLPAHGR 1888
Db 1663 NSSVTVVHATKSGPENPGSACRAWAYSAVNVYERDIHSLIGLPLICRKGILHMERNL 1722
QY 1889 QVTVOEALFTTIDETKSNVFTENMERNCRAPNIOVEDPTFKENRFRFAINGYIMDTL 1948
Db 1723 FMDMRPVLVFMVDEKSKWYEXS--KGRS-----RIESPEERNAHKFAVINGMYN-L 1774
QY 1949 PGLVMAQDQIRWLLSGNSNTHSHFSGHVFTVRKREYKMALNLYPGVETVEML 2008
Db 1775 PGLMYQEWRLHLLNMGSRDLHVHFGQTLDDRTKHQHLGWELLPGSKFTLEMK 1834
QY 2009 PSKAGIWRVECLIGEHLHAGMSTLFLVYNSKQCTPLGVNASGHIRDFOITASQYGQWAPK 2068
Db 1835 ASKPGWLLDTEVGENVAGVQWQTFPLIDCKCPMGLSTGVISDSQIKASEYLTYWEPR 1894
QY 2069 IARLHYSGSINAWTKE-----PRS-WIKDULLAPMIHGHKTKGAKQKPSLLISQPII 2122
Db 1895 LARLNNAGSYNAWIEKTALDFFPKPIQVDMQKVVVGTQGTQAKHYLKSCTFTFQV 1954
QY 2123 MYSLDGKKWYHRCNSTLTVFPGNVDSGSKENINFPPIIAEYVIRLHPHYHIRSTLR 2182
Db 1955 AYSDDQTNWQIFRGSKGSVMYFTGNSDGTIKENRLDPPVARYIRLHPYKSNRPTLR 2014
QY 2183 MELMGCLNSCSMPLGMESKAISDAQITASSYFTNMFAT-WSPSKARLHLQGRNWRPQ 2241
Db 2015 LELQGEVNGCSTPLGLEDRIGQIKQITASSFKKSGWGDYWEPSLARLNAQGRVNAWQAK 2074
QY 2242 VNNPKWLVDPFKMTKVTGVTQGVSKLLTSMYVKEPLISSQDGHOWTLFFQNGKV-- 2299
Db 2075 ANNNKQMLQVLLKTKKVTALVTOGCKSLSEMYKYSIYSGDQGVAWKPYROKSMVD 2134
QY 2300 KVFQGNQDSFTPVNSLDPPLLTLYRLRHPSQVWHQIALRMEVLGCEAQDLY 2351
Db 2135 KIFGNSNTKGMKNFPPIISRIRIIPKTNQSIARLLELPGC---DIY 2183
```

RESULT 6

```
KPB05
coagulation factor V precursor - bovine
C/Species: Bos primigenius taurus (cattle)
C/Date: 04-Mar-1993 #sequence_revision 28-Apr-1995 #text_change 11-Jun-1999
C/Accession: A42580; A36497
J/Guinito, E.R.; Esmon, C.T.; Mann, K.G.; MacGillivray, R.T.
J. Biol. Chem. 267, 2971-2978, 1992
A/Title: The complete cDNA sequence of bovine coagulation factor V.
A/Reference number: A42580; MUID:92147638; PMID:1737753
A/Accession: A42580
A/Molecule type: mRNA
A/Residues: 1-2211 <GUI>
```

```
A/Cross-references: GB:M81440; NID:g163037; PIDN:AAA30512.1; PID:g163038
A/Note: sequence extracted from NCBI backbone (NCBIN:80774, NCBI:P:80776)
R/Kalafatis, M.; Jenny, R.J.; Mann, K.G.
J. Biol. Chem. 265, 21580-21589, 1990
A/Title: Identification and characterization of a phospholipid-binding site of bovine fa
A/Reference number: A36497; MUID:91072354; PMID:2254316
A/Accession: A36497
A/Molecule type: protein
A/Residues: 1566-1570, 'X', 1572-1581, 'X', 1583-1584, 1673-1676, 'X', 1678-1679, 'X', 1681, 'X', 1
Biochem. J.; Kalafatis, M.; Silveira, J.R.; Kung, C.; Mann, K.G.
Biochemistry 33, 13109-13116, 1994
A/Title: Determination of the disulfide bridges in factor Va heavy chain.
A/Reference number: A55979; MUID:95034740; PMID:7947716
A/Contents: annotation
A/Note: 566-Cys and 617-Cys were shown to have free sulfhydryls
C/Comment: Factor V is activated by thrombin and partially by coagulation factor Xa.
C/Function:
A/Description: acts as a cofactor, with calcium and phospholipid, for the factor Xa prot
A/Pathway: blood coagulation
C/Superfamily: coagulation factor V; discoidin I amino-terminal homology; ferroxidase re
C/Keywords: blood coagulation; duplication; glycoprotein; phospholipid binding; plasma;
F/1-29/Domain: signal sequence #status predicted <SIG>
F/29-2211/Product: coagulation factor V #status predicted <MAT>
F/29-741/Product: coagulation factor Va heavy chain #status predicted <VAH>
F/29-345/Domain: A1 <DA1>
F/33-329/Domain: ferroxidase repeat homology <FO1>
F/346-695/Domain: A2 <DA2>
F/351-688/Domain: ferroxidase repeat homology <FO2>
F/656-1564/Domain: B <DOB>
F/1175-1437/Region: 9-residue repeats (Q-X-T/N-L-S-P-D-L-S)
F/1565-2211/Product: coagulation factor Va light chain #status predicted <VAL>
F/1565-1892/Domain: A3 <DA3>
F/1572-1892/Domain: ferroxidase repeat homology <FO3>
F/1854-1752/Region: phospholipid binding #status predicted
F/1883-2051/Domain: C1 <DC1>
F/1893-2048/Domain: discoidin I amino-terminal homology <DN1>
F/2052-2211/Domain: C2 <DC2>
F/2052-2208/Domain: discoidin I amino-terminal homology <DN2>
F/167-193, 248-329, 499-525/Disulfide bonds: #status experimental
F/225, 329, 397, 392, 460, 553, 587, 745, 756, 774, 780, 902, 952, 964, 1044, 1053, 1062, 1071, 1078, 1094,
F/334-335/Cleavage site: Arg-Asn (protein C) #status predicted
F/363, 657, 1537/Binding site: sulfate (Tyr) (covalent) #status predicted
F/376-377/Cleavage site: Arg-Ser (coagulation factor Xa) #status predicted
F/533-534/Cleavage site: Arg-Gly (protein C) #status predicted
F/607-688, 1712-1738, 1894-2048, 2053-2208/Disulfide bonds: #status predicted
F/741-742/Cleavage site: Arg-Ser (coagulation factor Xa, thrombin) #status predicted
F/1034-1035/Cleavage site: Arg-Ser (coagulation factor Xa, thrombin) #status predicted
F/1564-1565/Cleavage site: Arg-Ser (thrombin) #status experimental
```

```
Query Match 22.14; Score 2747; DB 1; Length 2211;
Best Local Similarity 30.24; Pred. No. 2,3e-135;
Matches 765; Conservative 382; Mismatches 833; Indels 550; Gaps 80;
```

```
QY 22 RRYLGAVELSWDMQSDGLPVDARPPRPVPSFPFNTSVVYKTLFVFTDHLFNIA 81
Db 32 RQFVAAQSRWNYR-----PESTHL-----SSKPFETS--PKKIVIRY-BAYPKE 76
QY 82 KPRPFWMLGPTTQAEVYDVTWTLKNMASHPVSLHAGVSVWKASGEAYDDQTSORE 141
Db 77 KQSRTSGLGPTLYAEVDIMKVHFNKHAHPLSHQAQIKYKFKSEGASYSHTLPM 136
QY 142 KEDDKVPFGSHTYVWQVLKENGPMASDPGLCLTYSLSHVDLVKDLNSGLIGALLVREG 201
Db 137 KMDDAVAPGQYTYEVIISHSHTDPPCLTHIYYSVNLVEDFNSGLIGPLICKKG 196
QY 202 SLAEKTKTLL--HKFILLFAVFDGKSWHSETKNSLMQDRDAASARAWPKHTYNGVNR 259
Db 197 TLTEDTQKPEKQCHVLMFAVDFESKSNQTS-----LMYTVNGYNG 240
QY 260 SLPLGLICHRKSVVWHVIGMGTTPPEVHSIFUEGHTFLVRNHRQASLEISPTFTTAQTL 319
Db 241 TMPDITVCAHDEISWHLIGMSSGPELFIHFNQGVLEONHHKISAITLVSAISTANMTV 300
```



```

Db 2145 SDQCTDWKPYREKSMYDKIFEGNNVRGHVKNFNFPIISRFIRIPKTIWNQSIARLE 2204
Qy 2342 VLGCERADLY 2351
Db 2205 LFQGC--DMY 22:1

RESULT 7
A25945
coagulation factor VIII - pig (fragment)
C;Species: Sus scrofa domestica (domestic pig)
C;Date: 16-Aug-1988 #sequence_revision 16-Aug-1988 #text_change 21-Jul-2000
C;Accession: A25945
R;Toole, J.J.; Pittman, D.D.; Orr, E.C.; Murcha, P.; Wasley, L.C.; Kaufman, R.J.
Proc. Natl. Acad. Sci. U.S.A. 83, 5939-5942, 1986
A;Title: A large region (approx 95 kba) of human factor VIII is dispensable for in vitro
A;Reference number: A25945; MUID:86287369; PMID:3016730
A;Accession: A25945
A;Status: nucleic acid sequence not shown; not compared with conceptual translation
A;Molecule type: DNA
A;Residues: 1-869 <TOO>
C;Superfamily: coagulation factor VIII; disocoidin I amino-terminal homology; ferroxidase

Query Match 17.6%; Score 2190; DB 2; Length 869;
Best Local Similarity 45.2%; Pred. No. 7.8e-107;
Matches 493; Conservative 122; Mismatches 251; Indels 224; Gaps 15;

Qy 705 GLWILGCHNSDFRNGRTALLKVVSCDKNYDSDYEDISAVLLSKNAIEPRSPSON 764
Db 1 GLWILGCHNSDLNRGMTALLKVVSCDKNYDSDYEDIPGFLSKNVIEPRSPSON 60

765 SRHPSSTOKQFNATVTPENDIEKTDPTFAHRTMPKIQNVSSDILLMLRQSPHGLSL 824
Db 61 SRPSSASQKQQTITSPDDVE-LDPOQERTQALELSVPSGDSMLLQNPAPHGSS 119

825 SDLOAKYETSDPSGAIDSNLSSEMTFRPOLHSGMVFTFPGSLQRLNEXLGT 884
Db 120 SDLOQEARNE--ADDYLPAGARENTPASAAARLPBLHHSABRLVTPPE--EK--- 167

885 TAAYELKGLDKVSTSTNNLS--TIPSDNLAAGTDNTSSIGPPSPVHYDSOLDITLFG 942
Db 168 ----BLKKLDSKMSSSDLTKTSTPTSDTSAETHTSLGPRHPQVNFRLSGALVIG 223

943 KKSSPLTSGGLPSSENNKSLIESGLMNSQSSWGVSTSGRLFKGKRAHGPA 1002
Db 224 KNSHFIAGVPLGSTEEDH-----ESSLGENVSPVSDGIFEKERAGFAS 270

1003 LTKNALPKVSIILTKNTKNSATNKTATKTHIDGSPSLIENSPPVQWQILSDTFKKVT 1062
Db 271 LTKDVLFPKVNISLVKTKARVYLKTNRKIHIDDAALLTENRASA----- 315

1063 PLIHDRMLMDKATALRLNHNKNTTSSKNMEMVQKKEGPTPPDPAQNPFMRMLFLP 1122
Db 316 -----TFMDKNTTASGLNHNVS----- 332

1123 BSARWIQTHGKNSLNSGGPKQKVLISLGPBKSVGGQNFLLSEKNKVVVGKEFTKDVGL 1182
Db 333 ---WKGLPKGNPLSSREGPSPELLTSSGSGSVKVGSGQGRIRVAVEBELSKG--- 385

1183 KMVFPSSRNPLPLNLDNLHNNNTINQEKIIOEEIEKKEITLIQENNVLPQIHVTGTGNF 1242
Db 386 KEMLPNSBLTFLTNSADVQGNTHSQGKKSEEMERREKLQVKVLDLPQVVTATGTKNF 445

1243 MKNLFLTSFRQVVEGSDGAYAPVLQDFPSINDSTNRTKHTAHPSKKEEENLEGLNQ 1302
Db 446 LRNIHQSTEPSVEGPDGGSHAPVPODSSLNDSAEATHLAHSAIREAPLRAPGNF 505

1303 TRQIVEKACTTRISNTSQCNFVTOQRKRLKQFLPLEETELEKRIIVDVTSTQWSKN 1362
Db 506 T-----GPCRSNVPVRVQSLKQLIELPLEIEIKPEGVVLNATSTWS-- 548

1363 MKHLTPSTLTQIDYNEKEKA-TQSPLSCLTRSHSIPQANKSPFLPIAKVSSFPSPRIPI 1422

```

```

Db 549 ----- 548
Qy 1423 LTRVLPQDNSSHLPAASVYRKKDSQVSSSHFLOGAKNNLSLAITLEMTGDQREVGSIG 1482
Db 549 -----ESSPILOAKNNLSLFLTLEMGAGGQKISALG 582
Qy 1483 TSATNSVYKKVENTVLPKFDLPKTSQKVELLPKVHIYOKDLFFPTETSGSPGHLDLVEG 1542
Db 583 KSAAGPLASGKLEKAVLSSAGLSEASCAEFLPKVRVREDLLLPQKTSNVSCHAGDLQCE 642
Qy 1543 SLQGTGALKAWEANRPGKVPFLRVATSSAKTPSKLLDPLAWDNHYGTQIPKEWKSQ 1602
Db 643 IFLOKTRGPNVLNKNVPPG-----RTPSKLLGP-----PMPK-EWESL 679
Qy 1603 EKSPKTAFAKKKDTI-LSLNACRSNHAIAINEQNKPEIEVTWAKQGRTERLCSQPPV 1661
Db 680 EKSPKSTALRTKDIISLPLDRHESNHSIAAKNEQATQREAAWTQGGPGRUGAPKPPV 739
Qy 1662 LKRQRHREITRTLLQSDQDEIDYDITISVMKKEDFDIYSDENOSPRSFKKTRHYFIAA 1721
Db 740 LRRHQDISLPTROPEDKMDYDDIFSTETKGEDFDIYSDENQDPRSFQKTRHYFIAA 799
Qy 1722 VERLWDYGMSSSPHVLNRAQSSVPOPKVVFQFTDGSFTQPLYSGLNEHLGLGPPY 1781
Db 800 VEQLWDYGMSSSPRALNRAQNGEVPFRKVVFRERADGSFTNPSYRGLNKHLLGLGPPY 859
Qy 1782 IRAEVEDNIM 1791
Db 860 IRAEVEDNIM 869

RESULT 8
KHU
ferroxidase (EC 1.16.3.1; precursor [validated] - human
N;Alternate names: ceruloplasmin
N;Contains: ferroxidase long form (CP-1); ferroxidase short form (CP-2)
C;Species: Homo sapiens (man)
C;References: 12-May-1995 #text_change 08-Dec-2000
C;Date: 31-Aug-1980 #sequence_revision 12-May-1995 #text_change 08-Dec-2000
C;Accession: A25443; A24165; A35450; A00524; I59067
R;Koschinsky, M.L.; Funk, W.D.; van Oost, B.A.; Macgillivray, R.T.A.
Proc. Natl. Acad. Sci. U.S.A. 83, 5086-5090, 1986
A;Title: Complete cDNA sequence of human ceruloplasmin.
A;Reference number: A25443; MUID:86259737; PMID:2873574
A;Accession: A25443
A;Molecule type: mRNA
A;Residues: 1-1060, 1065-1069 <KOS>
A;Cross-references: GB:M3689; NID:g180255; PIDN:AA451976.1; PID:g180256
A;Note: this is the short or CP-2 alternatively spliced form
R;Mercer, J.F.B.; Grimes, A.
FEBS Lett. 203, 185-190, 1986
A;Title: Isolation of a human ceruloplasmin cDNA clone that includes the N-terminal lead.
A;Reference number: A24165; MUID:86275241; PMID:3755405
A;Accession: A24165
A;Molecule type: mRNA
A;Residues: 1-40;549-599;784-829;919-952 <NER>
R;Yang, F.; Friedrichs, W.E.; Cupples, R.L.; Bonifacio, M.J.; Sanford, J.A.; Horton, W.A.
J. Biol. Chem. 265, 10780-10785, 1990
A;Title: Human ceruloplasmin. Tissue-specific expression of transcripts produced by alte.
A;Reference number: A35450; MUID:90285218; PMID:2355023
A;Accession: A35450
A;Molecule type: DNA
A;Residues: 1007-1064 <YAN>
A;Cross-references: GB:J05506
A;Note: this is the long or CP-1 alternatively spliced form
R;Takahashi, N.; Ortel, T.L.; Putnam, F.W.
Proc. Natl. Acad. Sci. U.S.A. 81, 390-394, 1984
A;Title: Single-chain structure of human ceruloplasmin: the complete amino acid sequence.
A;Reference number: A00524; MUID:84119493; PMID:6582496
A;Accession: A00524
A;Molecule type: protein
A;Residues: 20-1060, 1065-1069 <TAK>
A;Note: 79-Gly and 449-Gly were also found
R;Yang, F.; Naylor, S.L.; Lum, J.B.; Cutshaw, S.; McCombs, J.L.; Naberhaus, K.H.; McGill

```


Proc. Natl. Acad. Sci. U.S.A. 83, 3257-3261, 1986
A>Title: Characterization, mapping, and expression of the human ceruloplasmin gene.
A'Reference number: 159067; MUID:86205876; PMID:13456416
A'Accession: 159067
A>Status: translated from GB/EMBL/DBEJ
A'Molecule type: mRNA
A'Residuals: 218-1069 <RES>
A'Cross-references: GB:M13536; NID:g180248; PID:AAA51975.1; PID:g180249
C'Comment: Ferroxidase is a blue, plasma alpha2-glycoprotein binding 6-7 copper ions per molecule. In Wilson's disease the plasma levels of ferroxidase are diminished or undetectable. The three fragment chains are produced spontaneously during purification and processing.
C'Genetics:
A'Gene: GDB:CP
A'Cross-references: GDB:119069; OMIM:117700
A'Map position: 3q23-q25
A'Introns: 1066/3; 1061/1
A'Note: the list of introns is incomplete
C'Function:
A'Description: catalyzes the oxidation of free iron(II) to iron(III) coupled with the reduction of iron(III) to iron(II), is the form bound and transported by transferrin
A'Note: other possible functions are amine oxidase activity, copper transport and homeostasis
C'Superfamily: ferroxidase; ferroxidase repeat homology
C'Keywords: acute phase; alternative splicing; copper; duplication; glycoprotein; oxidoreductase
F'1-19/Domain: signal sequence #status predicted <SIG>
F'20-1069/Product: ferroxidase, short form #status predicted <MATL>
F'20-1060/Product: ferroxidase, short form #status experimental <MAUS>
F'20-499/Product: ferroxidase 67K chain #status experimental <K67>
F'23-357/Domain: ferroxidase repeat homology <F01>
F'373-718/Domain: ferroxidase repeat homology <F02>
F'501-905/Product: ferroxidase 50K chain #status experimental <K50>
F'733-1059/Domain: ferroxidase repeat homology <F03>
F'907-1065/Product: ferroxidase 19K chain #status experimental <K19>
F'138,397,762/Binding site: carbohydrate (Asn) (covalent) #status experimental
F'174-200,276,357,534-560,637-718,874-900/Disulfide bonds: #status predicted
F'227,338,926/Binding site: carbohydrate (Asn) (covalent) #status absent
F'295,338,343/Binding site: copper (His, Cys, His) (type 1) #status predicted
F'358/Binding site: carbohydrate (Asn) (covalent) (partial) #status experimental
F'656,699,704,709/Binding site: copper (His, Cys, His, Met) (type 1) #status predicted
F'994,1040,1045,1050/Binding site: copper (His, Cys, His, Met) (type 1) #status predicted

Query Match 10.7%; Score 1326; DB 1; Length 1069;

Best Local Similarity 19.2%; Pred. No. 2.1e-61;

Matches 399; Conservative 204; Mismatches 419; Indels 1054; Gaps 29;

QY 5 LSTCFCLLRFCFSATRYILGAVELSDYMQSDIGE---LPVDARPPRPVPGKFPFNT 61
DB 6 LGIFLFCSTP-AWAKEHYHIGIETWDY-ASDHGKLLSVDEHSNIYLONGPDR 63
QY 62 SYVYKTLFVEETHLNFIAKPRPPMGLLGTIIQAEVYDVTVTITLKNMASHFVSLHVG 121
DB 64 GRLYKKALYLYQTDFTFTTIEKPVMLGFLGPIIKAEITGDKVYVHLKNLASRPYTFHSHG 123
QY 122 VSYWZASGAEYDDOTSQREKEDDKVFPGGSTYVYVQVLEKNGPMASDPLCLTYSLSHV 181
DB 124 ITTYKEHGAIPYDNTDTPQRAADDKVYQGEQTYMLLATEEQSPGEGCNCVTRIYHSHI 183
QY 182 DAVKDLNSGLICALLVCRGSLAKETQTL-HKFTLLFAVPDEGKSWHSE-----TKNS 234
DB 184 DAPKIASGLIGLICCKDSLCKEKHIDREFVYFVSDENFSWYLEDNIKTYCSEP 243
QY 235 LMQDRDAASARAPKMTYNGVYVNSLPGLIGCHRSKSVYVHVGMTTPEVHSIFLEHT 294
DB 244 EKVDKDNEDFQESNRYSVNGYTFGSLPGLSCAEDRVKMWFLPGMGNEVDVLAAPFHGA 303
QY 295 FLVRNHRQASLSPIFTFLTAQTLMDLQGLLSCHISHQHDGMEAYVYVDSCEPQL 354
DB 304 LTNKNYRIDTINLFPATLFDAYVAQNPCEWMLSCQNHLKAGLQAFQVQEC----- 357
QY 355 RMKNKEAEYDDDLTSEMVDVRFDDNSPSFIQRSSVAKKPKTKTWHTYIAAEEDWDY 414
DB 358 ---NKSSKQ-----NTRGKHVRH-----YVIAAEELIWNV 385
QY 415 APL-----VLAPDDRSYKSYQLNNGFQRIGRKYKVRFMAYTDETF---XTREATQ 462

DB 386 APSGIDIFTKENTAPGSDS--AVFPEQGTTRIGGSYKLVREYTDASTNRKERGPBE 443
QY 463 HESGILGPLYGVGVSTLLIIIPKQASRPNIYPHGI-----TDVRLY---SRRLP 511
DB 444 EHLGILGPIVAEVDGTVIRVTFFNKAGYPLSIEPIGVRFNKNNEGTYYSNYPQGRSVP 503
QY 512 KGVKHLKDFPILPGEIFKFKWTVTVEDGPTKSDPRCLTRYYSFVNMERDLASGLIGPLL 571
DB 504 PSASH-----VATEFTYVTVKVGPNADPVCLAKMYSAVDPTKDFITGLIGEMK 558
QY 572 ICYKESVDQRGNQIMSDKRNVLFSVDFENRSWYLTENIQRFLENPAGVQLEDPQASN 631
DB 559 ICKKGLSHANGROKVDKEFLPTVDFENESLLEDNIRMFTTAPQVQKEDDFOESN 618
QY 632 IMHSINGYVFDLSIQ-LSVCLHEVAYWYILSIGAQTDFLSVFFSGYTFKHMVYEDTLTLF 690
DB 619 KMTSMNGFMYGNQFGLTMCCKGDSVWYLFSGAGNEADVHGIFYSGNTYLMWGERDRTANLF 678
QY 691 PFGSETVFMGMENPGLWILGCHNSDFENRGMTALLKVSSCDKNTGYEDSYEDISAYLL 750
DB 679 PQSLTLKMPDTEGTENVECLTTHVTGCKQKYTVNQC----- 718
QY 751 SKNAIBPRFSQNSRHPSTRQKQFNATTIPENDIEKTDWFAHRTPNPKIQNVSSDLL 810
DB 719 ----- 718
QY 811 MLRQSTPHGLSLSLDQAEKYEFTSDPSPGATDNNNSLSEMTHERPQLHHSMDMVFTP 870
DB 719 --RQSE-----DSTFYLGERTYI----- 735
QY 871 ESGQLRLNEKLGTAATELKLDFKVSSTNNLISTPSDNLAAAGTNTSSLGPPMPV 930
DB 736 -----AAVEVE----- 742
QY 931 HYDSQDITTLFGKXSPLTSGGGLSLSEENNDKLLSGLMNSQESGWKNVSTESGR 990
DB 743 ----- 742
QY 991 LFKGRAHPALLTKDNALFKVISLTKNTKTSNNSATNRKTHIDGPSLLIENSPPVQN 1050
DB 743 -----WD- 744
QY 1051 ILESDETFKVTPLIHDRMLMKDNATLRLNHNMGKTTSSKNMEMVOOKKEGPIPPDAQN 1110
DB 745 ----- 744
QY 1111 PDMSFFMFLPESARWIORTHGKNSLNSGGGSPKQLVSLGPEKSVGQNFLEKKNV 1170
DB 745 ----- 744
QY 1171 VGKGEFTKDVGLKEMVFPSSRNFLTLNLDNLHNHNTHNQEKKIQETEKETLIQENVVL 1230
DB 745 ----- 744
QY 1231 PQIHTVGTCKNFMKNLFLLSLRQNVESYDQAYAPVLQDFRSLANDSTNRTKKHTAHFSKK 1290
DB 745 -----YSP----- 747
QY 1291 GEENLEGLGNQTKQIVEKYACTTRISPTSQQNFVTOESKRAUKQFPLPLEETELEKRI 1350
DB 748 -----OR----- 749
QY 1351 IVDDTSTOWSKNMKHLTSTLTQIDYNEKEKGAITQSPSLDCLTRSHSIPQANRSLPIA 1410
DB 750 -----ZWEKELHHL----- 758
QY 1411 KVSSPSPRIYLTRVLFPQDNSSLPAASRYKKOSGVQESSHFLQAGKKNLSAILTLE 1470
DB 759 -----EQNVSNALF----- 768
QY 1471 MTGDQREVSLGTSATNSVTYKQVNTVLPKEDLPKTSKVELLPKVHYKQDLPPPTETS 1530


```

Db      769 ----- 768
QY      1531 NSSPCHLDLVEGSLLOGTGGAIKWNEANRPGKVPFLRVATESSAKTPSKLLDPLANDNEY 1590
Db      769 ----- 768
QY      1591 GTQIPKEEMKSOEKSEKTAFAKKKOTILSLNACSNHAIANEGONKPBIEVTWAKQGR 1650
Db      769 ----- 768
QY      1651 TERLCSQNPVLKRHOREITRTTLOSDBEIDYDDTISVEMKEDFDIYDENQSPRSF 1710
Db      769 ----- 772
QY      1711 QKTRHYFTIANVERLWDYGMSSPHVLNRAQSGVSPQKVVQFQFTDGSFTQPLVRGE 1770
Db      773 ----- 800
QY      1771 LNEHLGLLGPYTRAEYEDNIMVTRNQASRPYSFYSLSIYEDDQDQGAEPKRNFKV--- 1827
Db      801 ESEHLGILQPLHADVDGKVIIFKNWATPEYIHA-----HGVTESSTVTPTL 850
QY      1828 PNETKTYFWKVOHMAFTKDEPCKAWAPSDVDLEKDVHSGILGILLVCHTWTLPAGH 1887
Db      851 PGETLYWVKIPERSGAGTSDSACIPWAYSYSTVDQVKDLYSGILGILLVCRRLKVFNP 910
QY      1888 RQVTVQERAFFTIFDETSKSWYFTENNERNCRAPCNIQMEDPTKENYRFAHNGVIMDT 1947
Db      911 RRKL--EPALLFLVFNENESWYLDNKTYSDBPEKVNKDDEBFISNKKHANGRMFGN 968
QY      1948 LPLGVMAQDQRIWYLLSGNSNENIHSIHESGHVFTVRKKEEYKMAXLYNLPVGFVTEM 2007
Db      969 LQGLTHVGVDEVWYLLMGNGNEIDLTHVHGHGSFQYKHGVSDDVDFDPCTYQILEM 1028
QY      2008 LPSKAGIWRVECLIGELHAGMGTFLVYSNKKQTP 2043
Db      1029 FPRTPGTLWLFCHVDHIHAGMETTVTLQNEGYR 1054

RESULT 9
A35210
ferrooxidase (BC 1.16.3.1) precursor - rat
N;Alternate names: ceruloplasmin
C;Species: Rattus norvegicus (Norway rat)
C;Date: 10-Sep-1999 #sequence revision 10-Sep-1999 #text_change 10-Sep-1999
C;Accession: A35210; A41753; A29564; S21692
R;Fleming, R.B.; Gitlin, J.D.
J. Biol. Chem. 265, 7701-7707, 1990
A;Title: Primary structure of rat ceruloplasmin and analysis of tissue-specific gene exp
A;Reference number: A35210; MUID:90237081; PMID:2332446
A;Status: not compared with conceptual translation
A;Molecule type: mRNA
A;Residues: 1-1059 <FL2>
A;Cross-references: GB:J05424
R;Fleming, R.B.; Gitlin, J.D.
J. Biol. Chem. 267, 479-486, 1992
A;Title: Structural and functional analysis of the 5'-flanking region of the rat cerulop
A;Reference number: A41753; MUID:92112697; PMID:1730611
A;Accession: A41753
A;Molecule type: DNA
A;Residues: 1-48 <FL2>
A;Note: the authors translated the codon GAA for residue 40 as Gly, GAA for residue 41 a
R;Aldred, A.R.; Grimes, A.; Schreiber, G.; Mercer, J.F.B.
J. Biol. Chem. 262, 2875-2878, 1987
A;Title: Rat ceruloplasmin. Molecular cloning and gene expression in liver, choroid plex
A;Reference number: A29564; MUID:87137545; PMID:3818625
A;Accession: A29564
A;Molecule type: mRNA
A;Residues: 'NSG', 215-216, 'Y', 218, 'FAT', 222, 'F', 224-226, 'E', 228, 'LL', 231, 'D', 233-235, 'RY
A;Experimental source: liver
A;Note: the authors translated the codon GCG for residue 60 as Gly and GTG for residue 1
R;Ryan, T.P.; Grover, T.A.; Aust, S.D.

```

```

Arch. Biochem. Biophys. 293, 1-8, 1992
A;Title: Rat ceruloplasmin: resistance to proteolysis and kinetic comparison with human c
A;Reference number: S21692; MUID:92117681; PMID:1531003
A;Accession: S21692
A;Molecule type: protein
A;Residues: 20-29, 'Q', 902-910 <RYA>
C;Superfamily: ferroxidase; ferroxidase repeat homology
C;Keywords: copper; glycoprotein; oxidoreductase; plasma
F;1-19/Domain: signal sequence #status predicted <SIG>
F;20-1059/Product: ferroxidase #status predicted <MAT>
F;23-356/Domain: ferroxidase repeat homology <FOA>
F;372-712/Domain: ferroxidase repeat homology <FO2>
F;727-1053/Domain: ferroxidase repeat homology <FO3>

```

Query Match 10.3%; Score 1276; DB 1; Length 1059;

Best Local Similarity 19.1%; Pred. No. 8.7e-59;

Matches 395; Conservative 196; Mismatches 424; Indels 1054; Gaps 29;

```

QY      5  LSTCPFLCLLRFCFSATRRYYLGAVELSDWYQ--SDLGEL--PVDARPPRVPKSPFFNTS 62
Db      6  LSALLFL--HSSLAWTRKHHYIGTEAVWDVYASGSEKELISVDTEQSNFYLRNGPDRIG 64
QY      63  VYKKTLFVEFTDHLNIAKPPPMGLLGPITQAEVYDVITLKNMASHVPUSHAAGV 122
Db      65  RYKKALYSEYTDGFTXTIDKPAWLGFLGPVIAEVDKVSVHVKNFASRPYTFHAHV 124
QY      123 SYWASGAEYDDQTSQREKDDKVPFGGSHYVVOVLXENGPMSDPLCLTYSLSHVD 182
Db      125 TYTKANEALYPDNTDFQADKLPPGGQYLVVLA--NEPSPGSGDSNCVTFIYHSHVD 183
QY      183 LVKDLNSGLIGALLVCBEGSLAKBKQTTL--HKPILLFAYFDEGKSWHSETKSLM----- 236
Db      184 APKDIASGLIGLILLCKGSLHKEENIDQBFVLMFVVDENLSWYLEDNIKTFGSEPE 243
QY      237 -QDRDAASARAPKMTVNGYVNRSLPGLGCHKRSVYWHVIGMGTTPVHSLFLGHTF 295
Db      244 KVDKNEDEQESNRMYSLNGYTFGSLPGLSMAEDRVKWLFGMGNEVDHSLFHHQAL 303
QY      296 LVNRHQASLEISPTFLTAQTLMDLGQFLASCHTSSHQHDGMEAYKVVDSPPEPQLR 355
Db      304 TSKVHTDINLFPATLIDVSMVAQNPGVWMLSCQNLHLKAGLQAPFQVRDC----- 356
QY      356 MKQNEBAEDYDDLDTSEMVDVVRFDNDSFSTQIRSVAKKPKTWHYIADEEWDVYA 415
Db      357 --NKSP--DDDIQKRHV-----RH-----YYIAAEETIWDYA 385
QY      416 P-----LVLAPDORSYKSYLNNGPQIRGRKKYKVRFWAYTDEF---KTRAI 461
Db      386 PSTDFTTGENITSLGSDGRVPFEQ---GATIGGSYKLVNTRYTDDSFTRKRGPD 441
QY      462 QHESGILGPLNGEVDTLIIIFKNOASRFYNIYPHGTIDVR---PLYSRRRLPKGVKHL 517
Db      442 EEHLGILGPVIMAEVGDILRVTFHNKQFPLSIQPMGVTRFKENEGTYG---PDGRSSK 498
QY      518 KDFPILPGEILPKYKWTIVVDGTSKSPCLIFYYYSFVNMRDLASGLIGLLICYKES 577
Db      499 QASHVAPKETFTYEWTPKEMGPTYADPCLSKMYTSGVLDTKDITGLIGPMKICKGS 558
QY      578 VDQRGNQIMSDKRNVLFSVDNRNSWYLTENIQRLNPAQVQLEDPFQASNIMHSIN 637
Db      559 LLADGQKQVDKEFYLPATVFDENESLLDDNINMTTAPENVVDKEDDEQESNKHSMN 618
QY      638 GYVFDLSIQ--LSVCLHEVAYWYILISGAQTDFLSVFFSGYTFKHKVYVEDLTLPFSGET 696
Db      619 GPMYGNLPLGNMGLGESIWWLFSAGNEADVHGIIYFSGNYLSKGBERRDTANLPKHSLT 678
QY      697 VFMSMENPGLWILGCHNSDRPNQMTALLKVSCKDKNYGYEDSYEDISAYLLSKNNAI 756
Db      679 LLMTPTDEGSFDEVELTDTDTYGMKQKYTNOC-----KQGFEDVT----- 720
QY      757 EPRSFQNSRHPSTRQKQFNATTPENDIEKTDPFWAHRTMPKIQNVSSDILLMLRQS 816
Db      721 ----- 720

```

QY 817 PTPHGLSLDQEAQYETPSDDPSGALDENNSISEMTHPRPQLHHSGDMVFTPESGIQL 876
Db 721 -----
QY 877 RLNEKLGTTAATTELKLDKFKVSS*SNKLISTIPSDNLAAGTND*SSLGPPSM*VHYDSQL 936
Db 721 -----
QY 937 DTTLFGKKSP*LTSGGSLSEENNDKLLSGIMNSQSSWGKNVSS*TESGRLFKQKR 996
Db 721 -----
QY 997 AHGPALLTKDNALFKVISLILKTNKTSNNSATNRKTHIDGPSLLIENSPPSVKQNI*LESDT 1056
Db 727 TYYIAA -----
QY 1057 BFKVTP*LIHDM*MLDKNAT*RLNHSNKT*SSKNMEMVQKKRGP*IPPDQNPDM*SMFF 1116
Db 733 -----
QY 1117 KMLF*PESAR*IQTHGKNSLNSGQSP*PKQLVSLGPEKSVGGQ*FLSEKNKVVVGKGEF 1176
Db 733 -----
QY 1177 TKD*VGLKEMV*PPSRN*FLTNLDLNFENNTHNQEKIQEEIEKKETL*IQENWVLP*QIHTV 1236
Db 733 -----
QY 1237 TGT*KNFMKN*FLLS*TRQNVESYG*AYAPVLQDFRSLNDS*NRKTKHTAHFSK*GEE*NL 1296
Db 741 -----
QY 1297 EBLGNQ*TKQIVEKYACT*RI*SPNTSQQNFV*QSRKALKQ*FLPLETELEK*II*VD*TS 1356
Db 742 -----
QY 1357 TQWS*KNM*GHL*TPS*TLTQIDYNEK*KGAITQSP*SLDCL*TRSH*SIPOANR*SP*PIAKVSS*PP 1416
Db 743 RDWEMELHHL -----
QY 1417 S*RIPLYL*TRVLFQDN*SSHL*PAASYRK*QDSGV*ESSH*FLOGAKKNLS*LAIFL*EMTG*QOR 1476
Db 753 -----
QY 1477 EYVGL*TSATN*VYK*VENTVLP*PKDLP*KTSGKVELLP*KVH*YOKDL*PPT*TSNG*SPCH 1536
Db 763 -----
QY 1537 LDLVEG*SL*LGQ*TEGAIK*WNEAP*GKVP*FLRVAT*ESSAK*TPSKLLD*PLAWDN*HYGT*QIPK 1596
Db 763 -----
QY 1597 BEWKSQEK*SP*ETAFKKKDTILSLNAC*SNHAI*AIN*EQNKPEIEV*TWAKQ*RT*ERLCS 1656
Db 763 -----
QY 1657 QNPPVLK*RHQ*REITR*TLQSD*QBEIDYDDT*ISVEMK*KE*DFDI*VEDENQ*SP*SFQKK*TRH 1716
Db 763 -----
QY 1717 YFIAA*VERLWDY*GMSS*PHVLRN*RAQSG*VPQFKKYV*QFSD*TGST*QPL*YRGE*LN*EHLG 1776
Db 767 FRIGS -----
QY 1777 L*GPYIAE*VEDNIM*TFRNQ*AR*PVSFY*SSLLISY*ESDQ*RGAE*PRK*FVK---PNETKT 1833
Db 801 ILGL*LIHAD*VGAKV*VFKVM*ATRPYS*IIHA-----HGVT*KS*STV*APT*LPGE*VRT 850
QY 1834 YFWKV*CHHMA*PTKDE*FDCKAWAY*FSDV*LEKOVHSG*LI*GL*PLVC---HNTLN*PA*HG*QV 1890
Db 851 YI*WQIPERS*GAGTED*SP*CPWAY*ISIVDRVKDLIS*GLIG*PLIVCR*KS*YV*KN*FVK----- 905

QY 1891 TVQEFALFTIPDET*SKSWFTENMERN*CRAPCNIOMED*PTFKENYR*FHAINGYIMDTLPG 1950
Db 506 KNEFSL*FLVFDENES*WLD*DNIN*YFDHPEK*VKNKDNEEFISNKKHAI*NGKMG*FNLQ 965
QY -951 LVMAQ*QORIDWY*LLSGNS*ENIHSI*HPSGHV*FTRKKE*YKMALYN*LYPGVF*FTVEM*LP 2010
Db 966 L*TEHVG*DEVN*WYVAM*GNEID*LT*VHFHGS*FQKH*RGHIS*SDVDFDFP*GYOTLE*MPQ 1025
QY 2011 KAGIWR*VECLIG*EHLHAG*MT*FLV*VSNK 2039
Db 1026 TPGT*WLLH*CHVT*DH*HAG*MT*TV*VLPNQ 1054
RESULT 10
A44258
factor VIII-associated gene B hypothetical protein - human
C:Species: Homo sapiens (man)
C>Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 18-Jun-1999
C:Accession: A44258
R:Levinson, B.; Kenwright, S.; Camel, P.; Fisher, K.; Gitschier, J.
Genomics 14, 585-589, 1992
A:Title: Evidence for a third transcript from the human factor VIII gene.
A:Reference number: A44258; MUID:93052386; PMID:1427887
A:Accession: A44258
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-216 <LEV>
A:Cross-references: Q5:W90707; NID:G182316; PID:AAA58466.1; PID:G182317
C:Superfamily: coagulation factor VIII; discoidin I amino-terminal homology; ferroxidase
F:i-53/Domain: discoidin I amino-terminal homology #status atypical <DN1>
F:57-210/Domain: discoidin I amino-terminal homology <DN2>
Query Match 8.9%; Score 1104; DB 2; Length 216;
Best Local Similarity 100.0%; Pred. No. 6.7e-51;
Matches 208; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 2144 VFGNVDSG*GKHN*FPPII*ARVIRL*PHYSIR*STIR*LMELMG*CDLNS*CSMP*LGESKA 2203
Db 9 VFGNVDSG*GKHN*FPPII*ARVIRL*PHYSIR*STIR*LMELMG*CDLNS*CSMP*LGESKA 68
QY 2204 ISDAQITASSY*TNFATW*SPSKARL*H*QGRSNAR*POVNN*PKWLQYDFOK*TKV*TVGT 2263
Db 69 ISDAQITASSY*TNFATW*SPSKARL*H*QGRSNAR*POVNN*PKWLQYDFOK*TKV*TVGT 128
QY 2264 TQGVKSL*ITS*MYK*EFLISSQDGHQ*WTL*FTQNGK*VKV*FQGNQD*SF*TPVNSL*DP*PL*TR 2323
Db 129 TQGVKSL*ITS*MYK*EFLISSQDGHQ*WTL*FTQNGK*VKV*FQGNQD*SF*TPVNSL*DP*PL*TR 188
QY 2324 YLRH*PQSW*VHQIALR*MEVL*GCEA*QDLY 2351
Db 189 YLRH*PQSW*VHQIALR*MEVL*GCEA*QDLY 216
RESULT 11
JC4915
ags protein precursor - rat
N:Alternate names: O-acetyl-Cd3 ganglioside
C:Species: Rattus norvegicus (Norway rat)
C>Date: 26-Sep-1996 #sequence_revision 01-Nov-1996 #text_change 20-Jun-2000
C:Accession: JC4915
R:Ogura, K.; Nara, K.; Watanabe, Y.; Kohno, K.; Tai, T.; Sanai, Y.
Biochem. Biophys. Res. Commun. 225, 932-938, 1996
A:Title: Cloning and expression of cDNA for O-acetylation of GD3 ganglioside.
A:Cross-references: DBJ:U84068; NID:G1620006; PIDN:BAAL2210.2; PID:G1620007
A:Experimental source: CST cell
C:Comment: This protein is required for the O-acetylation of disialoganglioside sialic acid.
C:Genetics:
A:Gene: ags

```

QY      2209 ITASSYET-NMFA-TWSPSKARLHLQGRSNAWRPQVNNPKEWLQVDFOKTKKVGTITQ 2265
        :|::||: |::||: |::||: |::||: |::||: |::||: |::||: |::||: ||
Db      324 MGASSSYKTWNIRAFGWYPHLGELDNQGGINAWTAQSNSAKWLQVDLTQRQVTGIITQ 383
        :|::||: |::||: |::||: |::||: |::||: |::||: |::||: |::||: ||
QY      2266 GVKASLLTSMYYVEFLISSODGEHWTLFPONGKKVKVFQGNQDSFTPVVNSLDPPILLTRYL 2325
        :|::||: |::||: |::||: |::||: |::||: |::||: |::||: |::||: ||
Db      384 GARDPGHIVQVSRYKAHSDDGQWTIYEQQSSKFQGNLDNNGHKNIPEKPPIKRV 443
        :|::||: |::||: |::||: |::||: |::||: |::||: |::||: |::||: ||
QY      2326 RIHPQSWKHQIALRMVELGC 2345
        :|::||: |::||: |::||: |::||: |::||: |::||: |::||: |::||: ||
Db      444 RVLEFVSHNRITLRLELCG 463
        :|::||: |::||: |::||: |::||: |::||: |::||: |::||: |::||: ||

RESULT 13
Tl1743
DP47 protein- pig (fragment)
C;Species: Sus scrofa domestica (domestic pig)
C;Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 08-Sep-2002
C;Accession: Tl1743
R;Ensslin, M.; Vogel, T.; Calvete, J.J.; Thole, H.F.; Schmidtke, J.; Matsuda, T.; Toepfe
Bio1. Reprod. 58, 1057-1064, 1998
A;Title: Molecular cloning and characterization of P47, a novel boar sperm-associated zo
A;Reference number: Z17325; UID:98206817; PMID:9546740
A;Accession: Tl1743
A;Status: preliminary; translated from GB/EWSL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-409 <ENS>
A;Cross-references: EMBL:Y11683; NID:G2652927; PIDN:CAA72379.1; PID:G2652928
A;Experimental source: testis
C;Function:
A;Description: may be involved in membrane remodeling and/or function as a zona pellucid
C;Superfamily: milk fat globule protein; discoidin I amino-terminal homology; EGF homolog
F:6-40/Domain: egs homolog >EGF

```

Query Match. 5.2%; Score 650; DB 2; Length 409;
Best Local Similarity 39.3%; Pred. NO. 1.2e-26;
Matches 139; Conservative 62; Mismatches 125; Indels 28; Gaps 6;

```
QY 2018 ECLIGHLHAGMSTLFLVYSNK-----CQPLGMASGHIRDPOQTASQY-- 2062
DB 58 ECEVIDDAHRG--DVTEVICKCPHGTYGHCETICNAFLGMETGAIDPQISASMHLG 115
QY 2063 -----GQWAPKLARLHYSGSINAW-STKEPFGWIKVDLLAPMIIRGIKTQGARQRPSSY 2116
DB 116 FMGLQRAWELARLHRAGIVNAWTASNYDRNPWIOVNLIRMRVTVGVTVQASRAGASY 175
QY 2117 ISOFTIMYS-DGKKWQTYSGNSTGLTWLFVGNVDSSIKHNIENPPIIARYIRLHPHYS 2176
DB 176 MKTFKVAYSTDRKGFQFOIGARESGDKIFMGINDNSGLKVNLFEPVLEYVRLVPILCH 235
QY 2177 IRSTRLMGLGDLNCSMPLGMESKAISDAOITASSYFTN---MFATWSPSKARLHOG 2233
DB 236 RGCTLRFELLGCBELSCAPFLGLKDNTPNKQITASSFYRTWGLGAFSWPYFARLDNQG 295
QY 2234 RSNARWPOVNNPKNLOVPQKMTKVTGVTQGVKSLLTSMVYKFLISSSDGQHWITLF 2293
DB 296 KFNWTAQNSASAEWLQIDLGSGRRVTGIIITQCARDFGHICVYAAKVAISDDGVSWTET 355
QY 2294 FQNGKY--KVFGQNDQSFYVNVNSLDPLLIYLRIRHPSQSVHQIALRNEVLGC 2345
DB 356 RDQALEGKFFQNLNDNSHKKNMETPTLTFVRLFPVAWNIRTLRVLLGC 409
```

RESULT 14
S65138 glycoprotein antigen MGp57/53, mammary gland - bovine (fragment)
N;Alternate names: glycoprotein component 16/major fat-globule membrane protein/MFG-E8
C;Species: Bos primigenius taurus (cattle)
C;Date: 28-Oct-1996 #sequence_revision 13-Mar-1997 #text_change 07-Aug-1998
C;Accession: S65138; G48394
R;Aoki, N.; Kishi, M.; Taniguchi, Y.; Adachi, T.; Nakamura, R.; Matsuda, T.
Biochim. Biophys. Acta 1245, 385-391, 1995

A>Title: Molecular cloning of glycoprotein antigens MGP57/53 recognized by monoclonal an
 A:Reference number: S65138; MUID:96125736; PMID:8541316
 A:Accession: S65138
 A>Status: preliminary; not compared with conceptual translation
 A:Molecule type: mRNA
 A:Residues: 1-401 <AOK>
 R:Mather, I.H.; Banghart, L.R.; Lane, W.S.
 Biochem. Mol. Biol. Int. 29, 545-554, 1993
 A>Title: The major fat-globule membrane proteins, bovine components 15/16 and guinea-pig
 II-like sequences.
 A:Reference number: A48394; MUID:93250576; PMID:8485470
 A:Accession: G48394
 A>Status: preliminary
 A:Molecule type: protein
 A:Residues: 207-220 <WAT>
 A:Experimental source: milk
 A>Note: sequence extracted from NCBI backbone (NCBIP:131457)
 C:Superfamily: milk fat globule protein; discoidin I amino-terminal homology; EGF homolo
 C:Keywords: glycoprotein
 F:1-32/Domain: EGF homology <EG1>
 F:40-79/Domain: discoidin I amino-terminal homology <DN1>
 F:82-239/Domain: discoidin I amino-terminal homology <DN2>
 F:243-401/Domain: discoidin I amino-terminal homology <DN2>

Query Match 5.1%; Score 635; DB 2; Length 401;
 Best Local Similarity 37.9%; Pred. No. 7.1e-26;
 Matches 135; Conservative 66; Mismatches 123; Indels 32; Gaps 7;

QY 2018 ECLIGEHLAGMSTLFLVYSNK-----CQPLGMASCHIRDFQITAGQY-- 2062
 DB 50 EQVYDDSHRG--DVFIQVICKPLGVYGHCECTTCTSPGLMGTGAIADSQISASSMELG 107
 QY 2063 ----GQWAPKLARLHYSGSINAWST----KEPFSWIKVDLLAPMIHGKTCQARQKSS 2114
 DB 108 FNLGQWAPELARLHQTGIWNAWTSNGYDKNP--WIQVNLKMKWMTGVVTCGASRAGA 165
 QY 2115 LYISQFIIMYSLDGKKWQTYRGNSTGTLWVFFGNVDSGKIHFNPPILARYIRLHPH 2174
 DB 166 EYLKTFKVAYSTDGRQFQFIQVAGRSQDKIFIGNVNSGLKINLFDTPLEQYVPLVPII 225
 QY 2175 YSIRTLRMELMGCDLNSCMPLGMSKAIQITASSYFTN---MPATWSPKARLHL 2231
 DB 226 CHRGTCLRFELLGCELGCTEPLGKNTIPNKQITASSYKWTGLSAFSPFPYARLDN 285
 QY 2232 QGRSNAWRPOVNNKPEWLOVDFOKTMKVTGVTQGVKSLTSMYKFEFLISSQDGHQWT 2291
 DB 286 QGRNAWTAQNSASEWLOIDLGSKQKRVITGITQGRDFGHQYVAARVAYGDDGVWT 345
 QY 2292 LFFQNG--KVKVFGQNDSTFPVNSLDPPLLTRYLRHPQSWHQAIRMEVLGC 2345
 DB 346 EYKDPGASSEKIFPGNMNNSHKNIFFETPQARFVRIQPVAAHNRITLRVELLGC 401

RESULT 15
 S74211
 PAS-6/7 protein precursor - bovine
 N/Alternate names: glycoprotein component 16/major fat-globule membrane protein/MFG-E8 h
 C:Species: Bos primigenius taurus (cattle)
 C>Date: 04-Dec-1997 #sequence revision 12-Dec-1997 #text_change 04-Nov-2002
 C:Accession: S74211; S78114; S24181; S65138; G48394
 R:Harregarrd, J.; Andersen, M.H.; Berglund, L.; Rasmussen, J.T.; Petersen, T.E.
 Eur. J. Biochem. 240, 628-636, 1996
 A>Title: Characterization of glycoprotein PAS-6/7 from membranes of bovine milk fat glob
 A:Reference number: S74211; MUID:97008954; PMID:8956064
 A:Molecule type: mRNA
 A:Residues: 1-427 <HVA>
 A:Cross-references: EMBL:X91895; NID:g1632778; PIDN:CAA62997.1; PID:g1632779
 A:Accession: S78114
 A:Molecule type: protein
 A:Residues: 19-85;96-110;140-165;174-216;221-232;248-277;285-293;309-337;339-420;425-427
 R:Kim, D.H.; Kanno, C.; Mizokami, Y.
 Biochim. Biophys. Acta 1122, 203-212, 1992

A>Title: Purification and characterization of major glycoproteins, PAS-6 and PAS-7, from
 A:Reference number: S23926; MUID:92353107; PMID:1643094
 A:Accession: S24181
 A:Molecule type: protein
 A:Residues: 383-394 <KIM>
 R:Aoki, N.; Kishi, M.; Taniguchi, Y.; Adachi, T.; Nakamura, R.; Matsuda, T.
 Biochim. Biophys. Acta 1245, 385-391, 1995
 A>Title: Molecular cloning of glycoprotein antigens MGP57/53 recognized by monoclonal an
 A:Reference number: S65138; MUID:96125736; PMID:8541316
 A:Accession: S65138
 A>Status: preliminary; not compared with conceptual translation
 A:Molecule type: mRNA
 A:Residues: 27-427 <AOK>
 R:Mather, I.H.; Banghart, L.R.; Lane, W.S.
 Biochem. Mol. Biol. Int. 29, 545-554, 1993
 A>Title: The major fat-globule membrane proteins, bovine components 15/16 and guinea-pig
 II-like sequences.
 A:Reference number: A48394; MUID:93250576; PMID:8485470
 A:Accession: G48394
 A>Status: preliminary
 A:Molecule type: protein
 A:Residues: 233-246 <WAT>
 A:Experimental source: milk
 A>Note: sequence extracted from NCBI backbone (NCBIP:131457)
 C:Superfamily: milk fat globule protein; discoidin I amino-terminal homology; EGF homolo
 C:Keywords: blocked amino end; disulfide bond; glycoprotein; milk
 F:1-18/Domain: signal sequence #status predicted <SIG>
 F:19-427/Product: PAS-6/7 protein #status experimental <MAT>
 F:24-58/Domain: EGF homology <EG1>
 F:66-105/Domain: EGF homology <EG2>
 F:108-265/Domain: discoidin I amino-terminal homology <DN1>
 F:269-427/Domain: discoidin I amino-terminal homology <DN2>
 F:24-35;29-47;49-58;66-77;71-94;96-105/Disulfide bonds: #status predicted
 F:27/Binding site: carbohydrate (Ser) (covalent) #status experimental
 F:34/Binding site: carbohydrate (Thr) (covalent) #status experimental
 F:59;227/Binding site: carbohydrate (Asn) (covalent) #status experimental
 F:109-265;252-256;270-427/Disulfide bonds: #status experimental

Query Match 5.1%; Score 635; DB 2; Length 427;
 Best Local Similarity 37.9%; Pred. No. 7.9e-26;
 Matches 135; Conservative 66; Mismatches 123; Indels 32; Gaps 7;

QY 2018 ECLIGEHLAGMSTLFLVYSNK-----CQPLGMASCHIRDFQITAGQY-- 2062
 DB 76 EQVYDDSHRG--DVFIQVICKPLGVYGHCECTTCTSPGLMGTGAIADSQISASSMELG 133
 QY 2063 ----GQWAPKLARLHYSGSINAWST----KEPFSWIKVDLLAPMIHGKTCQARQKSS 2114
 DB 134 FNLGQWAPELARLHQTGIWNAWTSNGYDKNP--WIQVNLKMKWMTGVVTCGASRAGA 191
 QY 2115 LYISQFIIMYSLDGKKWQTYRGNSTGTLWVFFGNVDSGKIHFNPPILARYIRLHPH 2174
 DB 192 EYLKTFKVAYSTDGRQFQFIQVAGRSQDKIFIGNVNSGLKINLFDTPLEQYVPLVPII 251
 QY 2175 YSIRTLRMELMGCDLNSCMPLGMSKAIQITASSYFTN---MPATWSPKARLHL 2231
 DB 252 CHRGTCLRFELLGCELGCTEPLGKNTIPNKQITASSYKWTGLSAFSPFPYARLDN 311
 QY 2232 QGRSNAWRPOVNNKPEWLOVDFOKTMKVTGVTQGVKSLTSMYKFEFLISSQDGHQWT 2291
 DB 312 QGRNAWTAQNSASEWLOIDLGSKQKRVITGITQGRDFGHQYVAARVAYGDDGVWT 371
 QY 2292 LFFQNG--KVKVFGQNDSTFPVNSLDPPLLTRYLRHPQSWHQAIRMEVLGC 2345
 DB 372 EYKDPGASSEKIFPGNMNNSHKNIFFETPQARFVRIQPVAAHNRITLRVELLGC 427

Search completed: April 13, 2004, 14:09:38

Job time : 44 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 13, 2004, 14:01:27 ; Search time 18 seconds
(without alignments)
6800.941 Million cell updates/sec

Title: NP000123-328

Perfect score: 12416

Sequence: 1 MQIELSTCFCLLRFCFSA.....VVHQIALRVEVLGCEAQLDY 2351

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 segs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_42.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	12410	100.0	2351	1	FA8 HUMAN
2	8827	71.1	2319	1	FA8 MOUSE
3	8068	65.0	2133	1	FA8 PIG
4	2822.5	22.7	2224	1	FA5 HUMAN
5	2780.5	22.4	2258	1	FA5 PIG
6	2747	22.1	2211	1	FA5 BOVIN
7	1322	10.6	1065	1	CERU HUMAN
8	1273	10.3	1059	1	CERU RAT
9	1242	10.0	1062	1	CERU MOUSE
10	663	5.3	427	1	MFGM RAT
11	657	5.3	463	1	MFGM MOUSE
12	650	5.2	409	1	MFGM PIG
13	635	5.1	427	1	MFGM BOVIN
14	588	4.7	387	1	MFGM HUMAN
15	469.5	3.8	931	1	NRP2 HUMAN
16	464.5	3.7	925	1	NRP2 RAT
17	462.5	3.7	931	1	NRP2 MOUSE
18	458.5	3.7	914	1	NRP1 CHICK
19	451.5	3.6	922	1	NRP1 RAT
20	446.5	3.6	923	1	NRP1 MOUSE
21	443	3.6	928	1	NRP1 XENLA
22	429.5	3.5	923	1	NRP1 HUMAN
23	306.5	2.5	3133	1	HMCT BOMMO
24	266	2.1	764	1	CPX2 MOUSE
25	261	2.1	756	1	CPX2 HUMAN
26	260.5	2.1	280	1	XLR1 FUGRU
27	247	2.0	224	1	XLR1 MOUSE
28	243	2.0	224	1	XLR1 HUMAN
29	242.5	2.0	3418	1	BRC2 HUMAN
30	235.5	1.9	1358	1	SIR4 YEAST
31	224.5	1.8	3329	1	BRC2 MOUSE
32	223	1.8	722	1	CPXN MOUSE
33	216.5	1.8	1928	1	MYSL YEAST

34	217	1.7	734	1	CPXM HUMAN	Q96sm3 homo sapien
35	217	1.7	1331	1	CTA2 HUMAN	Q9ohc6 homo sapien
36	217	1.7	1381	1	YB57 YEAST	P34216 saccharomyc
37	216.5	1.7	1271	1	Y388 MYCCE	P47580 mycoplasma
38	213	1.7	1420	1	APX XENLA	Q01613 xenopus lae
39	212	1.7	1957	1	SPOF SCHPO	Q10411 schizosacch
40	212	1.7	3924	1	ANK2 HUMAN	Q01484 homo sapien
41	211	1.7	1284	1	NEK4 DROME	Q94887 drosophila
42	209.5	1.7	1177	1	Y307 MYCCE	P47549 mycoplasma
43	209	1.7	2867	1	RBP2 PLAYB	Q00799 plasmodium
44	208	1.7	1310	1	CTA4 MOUSE	Q99p47 mus musculu
45	206	1.7	1308	1	CTA4 HUMAN	Q9c0a0 homo sapien

ALIGNMENTS

RESULT 1

FA8_HUMAN	STANDARD;	PRT;	2351	AA.
AC	P00451.			
DT	21-JUL-1986 (Rel. 01, Created)			
DT	21-JUL-1986 (Rel. 01, Last sequence update)			
DT	10-OCT-2003 (Rel. 42, Last annotation update)			
DE	Coagulation factor VIII precursor (Procoagulant component)			
DE	(Antihemophilic factor) (AHF).			
GN	F8 OR F8C.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
OX	NCBI TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=86081164; PubMed=3935400;			
RA	Truett M.A., Blacker R., Burke R.D., Caput D., Chu C., Dina D.,			
RA	Hartog K., Kuo C.H., Maslarz F.R., Merryweather J.P., Najarian R.,			
RA	Pachl C., Potter S.J., Puma J., Quiroga M., Rall L.B., Randolph A.,			
RA	Urdea M.S., Valenzuela P., Dahl H.-H.M., Favalaro J., Hansen J.,			
RA	Nordfang O., Ezban M.,			
RT	"Characterization of the polypeptide composition of human factor			
RT	VIII:C and the nucleotide sequence and expression of the human kidney			
RT	CDNA."			
RL	DNA 4:333-349 (1985).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=85061548; PubMed=6438526;			
RA	Wood W.I., Capon D.J., Simonsen C.C., Eaton D.L., Gitschier J.,			
RA	Keyt B., Seeburg P.H., Smith D.H., Hollingshead P., Wilson K.L.,			
RA	Delwart E., Tuddenham E.G.D., Vohar G.A., Lawn R.M.,			
RT	"Expression of active human factor VIII from recombinant DNA clones."			
RL	Nature 312:330-337 (1984).			
RN	[3]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=85061550; PubMed=6438528;			
RA	Toole J.J., Knopf J.L., Wozney J.M., Sultzman L.A., Buecker J.L.,			
RA	Pittman D.D., Kaufman R.J., Brown E., Shoemaker C., Orr B.C.,			
RA	Amphlett G.W., Foster W.B., Coe M.L., Knutson G.J., Pass D.N.,			
RA	Hewick R.M.,			
RT	"Molecular cloning of a cDNA encoding human antihemophilic factor."			
RL	Nature 312:342-347 (1984).			
RN	[4]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=93265012; PubMed=1303178;			
RA	Gitschier J., Wood W.I.,			
RT	"Sequence of the exon-containing regions of the human factor VIII			
RT	gene."			
RL	Hum. Mol. Genet. 1:199-200 (1992).			
RN	[5]			
RP	SEQUENCE OF 2064-2070 FROM N.A.			
RA	de Water N.S., Williams R., Browett P.J.,			
RL	Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.			
RN	[6]			
RP	SULFATION OF TYR-1699.			

RX MEDLINE=91093266; PubMed=1898735;
 RA Leyte A., van Schijndel H.B., Niehrs C., Huttner W.B., Verbeest M.P.,
 RA Mertens K., van Mourik J.A.;
 RT "Sulfation of Tyr1680 of human blood coagulation factor VIII is
 RT essential for the interaction of factor VIII with von Willebrand
 RT factor."; Chem. 266:740-746(1991).
 RL J. Biol. Chem. 266:740-746(1991).
 RN [7]
 RP SULFATION.
 RX MEDLINE=92207952; PubMed=1554716;
 RA Pittman D.D., Wang J.H., Kaufman R.J.;
 RT "Identification and functional importance of tyrosine sulfate
 RT residues within recombinant factor VIII."; Biochemistry 31:3315-3325(1992).
 RL Biochemistry 31:3315-3325(1992).
 RN [8]
 RP STRUCTURE BY NMR OF 2322-2343.
 RX MEDLINE=95200924; PubMed=7893714;
 RA Gilbert G.B., Baleja J.D.;
 RT "Membrane-binding peptide from the C2 domain of factor VIII forms an
 RT amphipathic structure as determined by NMR spectroscopy."; Biochemistry 34:3022-3031(1995).
 RL Biochemistry 34:3022-3031(1995).
 RN [9]
 RP REVIEW ON MOLECULAR BASIS OF HEMA.
 RX MEDLINE=91221499; PubMed=1902642;
 RA Gitschier J.;
 RT "The molecular basis of hemophilia A."; Ann. N.Y. Acad. Sci. 614:89-96(1991).
 RL Ann. N.Y. Acad. Sci. 614:89-96(1991).
 RN [10]
 RP REVIEW ON MOLECULAR BASIS OF HEMA.
 RX MEDLINE=89088506; PubMed=2491949;
 RA White G.C. II, Shoemaker C.B.;
 RT "Factor VIII gene and hemophilia A."; Blood 73:1-12(1989).
 RL Blood 73:1-12(1989).
 RN [11]
 RP REVIEW ON MOLECULAR BASIS OF HEMA.
 RX MEDLINE=95245332; PubMed=7728145;
 RA Antonarakis S.E., Kazazian H.H., Tuddenham E.G.D.;
 RT "Molecular etiology of factor VIII deficiency in hemophilia A."; Hum. Mutat. 5:1-22(1995).
 RL Hum. Mutat. 5:1-22(1995).
 RN [12]
 RP VARIANT HEMA GLN-2326.
 RX MEDLINE=86235434; PubMed=3012775;
 RA Gitschier J., Wood W.I., Shuman M.A., Lawn R.M.;
 RT "Identification of a missense mutation in the factor VIII gene of a
 RT mild hemophilic."; Science 232:1415-1416(1986).
 RL Science 232:1415-1416(1986).
 RN [13]
 RP VARIANT HEMA PRO-2135.
 RX MEDLINE=88096339; PubMed=3122181;
 RA Levinson B., Janco R.L., Phillips J.A. III, Gitschier J.;
 RT "A novel missense mutation in the factor VIII gene identified by
 RT analysis of amplified hemophilia DNA sequences."; Nucleic Acids Res. 15:9797-9805(1987).
 RL Nucleic Acids Res. 15:9797-9805(1987).
 RN [14]
 RP VARIANT HEMA GLN-2228.
 RX MEDLINE=88191889; PubMed=2893855;
 RA Youssoufian H., Antonarakis S.E., Bell W., Griffin A.M.,
 RA Kazazian H.H.;
 RT "Nonsense and missense mutations in hemophilia A: estimate of the
 RT relative mutation rate at CG dinucleotides."; Am. J. Hum. Genet. 42:718-725(1988).
 RL Am. J. Hum. Genet. 42:718-725(1988).
 RN [15]
 RP VARIANT HEMA GLY-291.
 RX MEDLINE=88220354; PubMed=2835904;
 RA Youssoufian H., Wong C., Aronis S., Platakouk H., Kazazian H.H. Jr.,
 RA Antonarakis S.E.;
 RT "Moderately severe hemophilia A resulting from Glu-->Gly substitution
 RT in exon 7 of the factor VIII gene."; Am. J. Hum. Genet. 42:867-871(1988).
 RL Am. J. Hum. Genet. 42:867-871(1988).
 RN [16]
 RP VARIANT HEMA CYS-1708.
 RX MEDLINE=89274393; PubMed=2499363;
 RA O'Brien D.P., Tuddenham E.G.;

RT "Purification and characterization of factor VIII 1,689-Cys: a
 RT nonfunctional cofactor occurring in a patient with severe hemophilia
 RL Blood 73:2117-2122(1989).
 RN [17]
 RP VARIANT HEMA CYS-391.
 RX MEDLINE=90001543; PubMed=2506948;
 RA Shima M., Ware J., Yoshioka A., Fukui H., Fulcher C.A.;
 RT "An arginine to cysteine amino acid substitution at a critical
 RT thrombin cleavage site in a dysfunctional factor VIII molecule."; Blood 74:1612-1617(1989).
 RL Blood 74:1612-1617(1989).
 RN [18]
 RP VARIANT HEMA LEU-189.
 RX MEDLINE=90057680; PubMed=2510835;
 RA Chan V., Chan T.K., Tong T.M., Todd D.;
 RT "A novel missense mutation in exon 4 of the factor VIII:C gene
 RT resulting in moderately severe hemophilia A."; Blood 74:2688-2691(1989).
 RL Blood 74:2688-2691(1989).
 RN [19]
 RP VARIANT HEMA LEU-2326.
 RX MEDLINE=89197216; PubMed=2495245;
 RA Inaba H., Fujimaki M., Kazazian H.H. Jr., Antonarakis S.E.;
 RT "Mild hemophilia A resulting from Arg-to-Leu substitution in exon 26
 RT of the factor VIII gene."; Hum. Genet. 81:335-338(1989).
 RL Hum. Genet. 81:335-338(1989).
 RN [20]
 RP VARIANT HEMA HIS-391.
 RX MEDLINE=89264602; PubMed=2498882;
 RA Arai M., Inaba H., Higuchi M., Antonarakis S.E., Kazazian H.H. Jr.,
 RA Fujimaki M., Hoyer L.W.;
 RT "Direct characterization of factor VIII in plasma: detection of a
 RT mutation altering a thrombin cleavage site (arginine-372-->histidine)."; Proc. Natl. Acad. Sci. U.S.A. 86:4277-4281(1989).
 RL Proc. Natl. Acad. Sci. U.S.A. 86:4277-4281(1989).
 RN [21]
 RP VARIANT HEMA CYS-1708.
 RX MEDLINE=90105723; PubMed=2104766;
 RA Arai M., Higuchi M., Antonarakis S.E., Kazazian H.H. Jr.,
 RA Phillips J.A. III, Janco R.L., Hoyer L.W.;
 RT "Characterization of a thrombin cleavage site mutation (Arg 1689 to
 RT Cys) in the factor VIII gene of two unrelated patients with
 RT cross-reacting material-positive hemophilia A."; Blood 75:384-389(1990).
 RL Blood 75:384-389(1990).
 RN [22]
 RP VARIANTS HEMA GLN-2228 AND LEU-2326.
 RX MEDLINE=90123183; PubMed=2105106;
 RA Casula L., Murru S., Pecorara M., Ristaldi M.S., Restagno G.,
 RA Mancuso G., Morfini M., de Biasi R., Baudo F., Carbonara A.;
 RT "Recurrent mutations and three novel rearrangements in the factor
 RT VIII gene of hemophilia A patients of Italian descent."; Blood 75:662-670(1990).
 RL Blood 75:662-670(1990).
 RN [23]
 RP VARIANT HEMA CYS-391.
 RX MEDLINE=90329422; PubMed=1973901;
 RA Pattinson J.K., McVey J.H., Boon M., Ajani A., Tuddenham E.G.;
 RT "CRM+ hemophilia A due to a missense mutation (372-->Cys) at the
 RT internal heavy chain thrombin cleavage site."; Br. J. Haematol. 75:73-77(1990).
 RL Br. J. Haematol. 75:73-77(1990).
 RN [24]
 RP VARIANTS HEMA PHE-1699 AND CYS-1708.
 RX MEDLINE=90152691; PubMed=2105906;
 RA Higuchi M., Wong C., Kochan L., Olek K., Aronis S., Kasper C.K.,
 RA Kazazian H.H., Antonarakis S.E.;
 RT "Characterization of mutations in the factor VIII gene by direct
 RT sequencing of amplified genomic DNA."; Genomics 6:65-71(1990).
 RL Genomics 6:65-71(1990).
 RN [25]
 RP VARIANTS HEMA CYS-1728 AND ASP-1941.
 RX MEDLINE=90169988; PubMed=2106480;
 RA Travstman M.D., Higuchi M., Kasper C.K., Antonarakis S.E.,
 RA Kazazian H.H.;
 RT "Use of denaturing gradient gel electrophoresis to detect point
 RT mutations in the factor VIII gene.";

Query Match			
Best Local Similarity 100.0%; Score 12410; DB 1; Length 2351;			
Matches 2350; Conservative 0; Mismatches 1; Indels 0; Gaps 0;			
Qy	1	MQIELSTCFELCLLRFCFSATRRYYLGAVE-SWDYMSQDLGELVDAFFPRVPKSPFFN	60
Db	1	MQIELSTCFELCLLRFCFSATRRYYLGAVELSWDYMSQDLGELVDAFFPRVPKSPFFN	60
Qy	61	TSVVYKKTLPVEFTDHLFNTAKRPPPMWGLLGPTIOAEVVDTVVITLKNMASHPVSLHAV	120
Db	61	TSVVYKKTLPVEFTDHLFNTAKRPPPMWGLLGPTIOAEVVDTVVITLKNMASHPVSLHAV	120
Qy	121	GVSWKASGAEYDDQTSQREKEDKVPFGGSHYVWQVLKENGPMASDPLCLTYSLSH	180
Db	121	GVSWKASGAEYDDQTSQREKEDKVPFGGSHYVWQVLKENGPMASDPLCLTYSLSH	180
Qy	181	VDLVKDLNSGLIGALLYCREGSLAKEKTQTLHKFILLFAVFDGKSWHSETKNSLMQDRD	240
Db	181	VDLVKDLNSGLIGALLYCREGSLAKEKTQTLHKFILLFAVFDGKSWHSETKNSLMQDRD	240
Qy	241	AASARAWPKMHTVNGYVNRSLPGLIGHRKSVYWHVIGMGTTPVHSEIFLEGHTFLVRNH	300
Db	241	AASARAWPKMHTVNGYVNRSLPGLIGHRKSVYWHVIGMGTTPVHSEIFLEGHTFLVRNH	300
Qy	301	ROASLEISPIITELTAQTLMDLGOFLLSCHLISSHQHDMBAVYKVDSCPBEPQLRMKNNE	360
Db	301	ROASLEISPIITELTAQTLMDLGOFLLSCHLISSHQHDMBAVYKVDSCPBEPQLRMKNNE	360
Qy	361	BAEDYDDDLTDEMDVVRFDNDNSPSFIQIRSAVAKHPKTWVHYIAAEBEEDWDYAPLVLA	420
Db	361	BAEDYDDDLTDEMDVVRFDNDNSPSFIQIRSAVAKHPKTWVHYIAAEBEEDWDYAPLVLA	420
Qy	421	PDSSYSQVLNNGPQIRGKYKVRMAVYDDEFKTRREALQHESGILGPLLYGEGVDTL	480
Db	421	PDSSYSQVLNNGPQIRGKYKVRMAVYDDEFKTRREALQHESGILGPLLYGEGVDTL	480
Qy	481	LILFKQASRPYNIYPHGIITDVRPLYRRPLKGVKHLKDPILPGEIPFKYKWTVTVEDGP	540
Db	481	LILFKQASRPYNIYPHGIITDVRPLYRRPLKGVKHLKDPILPGEIPFKYKWTVTVEDGP	540
Qy	541	TKSDPRCLTRYYSFVANNERDLASGLIGPLLI CYKESVDQRGNQIMSDKRNVLFSVFDE	600
Db	541	TKSDPRCLTRYYSFVANNERDLASGLIGPLLI CYKESVDQRGNQIMSDKRNVLFSVFDE	600
Qy	601	NRSWYLTENQRLPUNPAGVQLEDEPFOASNIMESINGYVFDLSQLSVCLHEVAYWYILS	660
Db	601	NRSWYLTENQRLPUNPAGVQLEDEPFOASNIMESINGYVFDLSQLSVCLHEVAYWYILS	660
Qy	661	IGAQTDFLSVFFSGYTFKHKMWYEDTLTLFPFSGETVFMSENPGILWLGCHNSDPNRNG	720
Db	661	IGAQTDFLSVFFSGYTFKHKMWYEDTLTLFPFSGETVFMSENPGILWLGCHNSDPNRNG	720
Qy	721	MTALLKVVSSCDKNTGDIYEDSYEDI SAKLSKNAIAEPSPQNSRHPSTKQKQFNATTI	780
Db	721	MTALLKVVSSCDKNTGDIYEDSYEDI SAKLSKNAIAEPSPQNSRHPSTKQKQFNATTI	780
Qy	781	PENDIEKTDKFAHRTDMPKIQNVSSDLMLLQSPSTPHGLSLSDIQAKEYETFSDDPS	840
Db	781	PENDIEKTDKFAHRTDMPKIQNVSSDLMLLQSPSTPHGLSLSDIQAKEYETFSDDPS	840
Qy	841	PGAIDSNNSISEMTHFRPOLHHSQDMVFTPSGQLRLNEKLGTTAATLKKLDFKVSST	900
Db	841	PGAIDSNNSISEMTHFRPOLHHSQDMVFTPSGQLRLNEKLGTTAATLKKLDFKVSST	900
Qy	901	SNNLISITIPSDNLAAQTDNTSSLGPPMPVHYDSQDLDTTLFGKKSPLTBSGGPLSLSEE	960
Db	901	SNNLISITIPSDNLAAQTDNTSSLGPPMPVHYDSQDLDTTLFGKKSPLTBSGGPLSLSEE	960
Qy	961	NNDSKLESGLMNSQESSGWQVSSTSRGLFKGKRAHGPAALLTKDNALPKVSIISLLKTN	1020
Db	961	NNDSKLESGLMNSQESSGWQVSSTSRGLFKGKRAHGPAALLTKDNALPKVSIISLLKTN	1020
Qy	1021	KTSNNSATNEXKTHIDGPSLLIENS SVWQNLISDTEFKKVTPLIHDRMLMDKNATLRL	1080
Db	1021	KTSNNSATNEXKTHIDGPSLLIENS SVWQNLISDTEFKKVTPLIHDRMLMDKNATLRL	1080
Qy	1081	NHXSNTKTSSKNMEMVQOKKEGPIPPDAQNPDASFFKXMLFLPDSARMIQRTGKNSLNSG	1140
Db	1081	NHXSNTKTSSKNMEMVQOKKEGPIPPDAQNPDASFFKXMLFLPDSARMIQRTGKNSLNSG	1140
Qy	1141	QGSPPKQVSLGPEKSVGEGNQFLSEKKNVVGKEFTKDVGLKEMVPPSSRNLFPLTNLDN	1200
Db	1141	QGSPPKQVSLGPEKSVGEGNQFLSEKKNVVGKEFTKDVGLKEMVPPSSRNLFPLTNLDN	1200
Qy	1201	LHENNTHNQKKQBEIEKKEKTELIOENVLPQIHTVTGCTKNMKNLFLLSRQNVVSGSYD	1260
Db	1201	LHENNTHNQKKQBEIEKKEKTELIOENVLPQIHTVTGCTKNMKNLFLLSRQNVVSGSYD	1260
Qy	1261	GAYAPVLQDFRSLNDSNTRTKKHTAHFSKKEBEENLEGLGNQTKQIVEKYACTRISPT	1320
Db	1261	GAYAPVLQDFRSLNDSNTRTKKHTAHFSKKEBEENLEGLGNQTKQIVEKYACTRISPT	1320
Qy	1321	SOQNFVTRQSKRALKQRLPIEBETELEKRIIVDDTSTQWSKNMKHLTPSTILTQIDYNEKE	1380
Db	1321	SOQNFVTRQSKRALKQRLPIEBETELEKRIIVDDTSTQWSKNMKHLTPSTILTQIDYNEKE	1380
Qy	1381	KGAIQTOSPLDCLTRSHSIPOANRSPUPIAKVSFSPSIRPIYLRVLFDQNSSHLPAASY	1440
Db	1381	KGAIQTOSPLDCLTRSHSIPOANRSPUPIAKVSFSPSIRPIYLRVLFDQNSSHLPAASY	1440
Qy	1441	RKDSGVQESSHFLQGAACKNNLSAILTLEMTGQREVGSIGTSATNSVTYKXVENTVLP	1500
Db	1441	RKDSGVQESSHFLQGAACKNNLSAILTLEMTGQREVGSIGTSATNSVTYKXVENTVLP	1500
Qy	1501	KPDLPKTSGKVELLPKVHIYQKDLFPPTETNSGSPGHLDLVEGSLLOGTBGAIKWNEANRP	1560
Db	1501	KPDLPKTSGKVELLPKVHIYQKDLFPPTETNSGSPGHLDLVEGSLLOGTBGAIKWNEANRP	1560
Qy	1561	GKVPFLRVATESAKTPSKLLDPLANDNHYGTQIPKEEWSQEKSPKTAFAKKKDTILSL	1620
Db	1561	GKVPFLRVATESAKTPSKLLDPLANDNHYGTQIPKEEWSQEKSPKTAFAKKKDTILSL	1620
Qy	1621	NACESNHAI AAINEGONKPEIEVTWAKQGTBELCSQNPVLPKKEHOREIIRTTLQSDOEE	1680
Db	1621	NACESNHAI AAINEGONKPEIEVTWAKQGTBELCSQNPVLPKKEHOREIIRTTLQSDOEE	1680
Qy	1681	IDYDDTISVEMKKEDFDIYDEENQSPRSQKTRHYFIAAVERLWDYGMSSSPHVLNR	1740
Db	1681	IDYDDTISVEMKKEDFDIYDEENQSPRSQKTRHYFIAAVERLWDYGMSSSPHVLNR	1740
Qy	1741	AQSGSVPOPKVVFQBEFTDGSFTQPLVRGELNHLGLLGPYIRAEVEDNTMVTFRNQASR	1800
Db	1741	AQSGSVPOPKVVFQBEFTDGSFTQPLVRGELNHLGLLGPYIRAEVEDNTMVTFRNQASR	1800
Qy	1801	PYSFYSSLISYEBEDQGAAPRKVPKNETKTYFWKVOHHMAPT KDEPCKAWAYPSDV	1860
Db	1801	PYSFYSSLISYEBEDQGAAPRKVPKNETKTYFWKVOHHMAPT KDEPCKAWAYPSDV	1860
Qy	1861	DLEKDVHSGHIGLBYLCHTNTINPAHQRTVQBEFALPFTIPBETKSWYETENMERNCR	1920
Db	1861	DLEKDVHSGHIGLBYLCHTNTINPAHQRTVQBEFALPFTIPBETKSWYETENMERNCR	1920
Qy	1921	PCNIQEMDPTFKENYFPAHNGVIMDTPLGLVMAQQRIRWYLLSMGSNENIESIHFSGH	1980
Db	1921	PCNIQEMDPTFKENYFPAHNGVIMDTPLGLVMAQQRIRWYLLSMGSNENIESIHFSGH	1980
Qy	1981	VFTVRKKEEYKMALYNLYPCGVFTVEMLPSKAGIWRVECLIGEHLHAGMSTLFLVYSNKC	2040
Db	1981	VFTVRKKEEYKMALYNLYPCGVFTVEMLPSKAGIWRVECLIGEHLHAGMSTLFLVYSNKC	2040
Qy	2041	QTPLGWASGHIRFOITASQYQGWAPKLARLHYSGSINAWSTKEPPSWIKVLLAPMII	2100
Db	2041	QTPLGWASGHIRFOITASQYQGWAPKLARLHYSGSINAWSTKEPPSWIKVLLAPMII	2100
Qy	2101	HGIKTQGARQKFSLSYISQFIIMYSLDGKKWQYRGNSTGTLMVFFGCVNDSGSKHNIFN	2160


```

Db 2101 HGKIQGARQKFSLLYSIOFIYSLDGGKQVYRGNSGTGLWFFGNGVDSGKIKHNIFN 2160
Qy 2161 PPIIARYIRLHPHTHSIRSTRLMELMGDLNCSNPLGMSKASISDAQITASSYPTNMPA 2220
Db 2162 PPIIARYIRLHPHTHSIRSTRLMELMGDLNCSNPLGMSKASISDAQITASSYPTNMPA 2220
Qy 2221 TWSPSKARLHLQGRNANRPPQNNPKEMQLQVDFQKTMKVTGVTGQVKSLLTSMYVKEFL 2280
Db 2221 TWSPSKARLHLQGRNANRPPQNNPKEMQLQVDFQKTMKVTGVTGQVKSLLTSMYVKEFL 2280
Qy 2281 ISSSQDGHQWTLFFQNGKVKVFGQNDSTPVVNSLDPELTRYRIHPQSWHQAIALRM 2340
Db 2281 ISSSQDGHQWTLFFQNGKVKVFGQNDSTPVVNSLDPELTRYRIHPQSWHQAIALRM 2340
Qy 2341 EVLGCEAQLLY 2351
Db 2341 EVLGCEAQLLY 2351

RESULT 2
FA8_MOUSE
ID FA8_MOUSE STANDARD; PRT; 2319 AA.
AC Q06194;
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
D2 Coagulation factor VIII precursor (Procoagulant component).
GN F8 OR CF8 OR F8C.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1] _
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6 X CBA; TISSUE=Liver;
RX MEDLINE=93300511; PubMed=8314577;
RA Elder B., Lakich D., Gitschier J.;
RT "Sequence of the murine factor VIII cDNA.";
RL Genomics 16:374-379(1993).
CC -!- FUNCTION: Factor VIII, along with calcium and phospholipid, acts
CC as a cofactor for factor IXa when it converts factor X to the
CC activated form, factor Xa.
CC -!- SUBCELLULAR LOCATION: Extracellular.
CC -!- TISSUE SPECIFICITY: Found in most tissues.
CC -!- SIMILARITY: Contains 3 F5/8 type A domains.
CC -!- SIMILARITY: Contains 2 F5/8 type C domains.
CC -!- SIMILARITY: STRONG, TO COAGULATION FACTOR V.
CC
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; L05573; AAA37385.1; -.
CC PIR; A47004; A47004.
CC HSSP; P00451; 1CFG.
CC MGD; MGI:88383; F8.
CC InterPro; IPR001117; Cu-oxidase.
CC InterPro; IPR008972; Cupredoxin.
CC InterPro; IPR000421; FA58 C.
CC InterPro; IPR008979; GalBind_like.
CC Pfam; PF00394; Cu-oxidase; 3.
CC Pfam; PF00754; F5_F8_type_C; 2.
CC SMART; SM00231; FA58C; 2.
CC PROSITE; PS01285; FA58C_1; 2.
CC PROSITE; PS01286; FA58C_2; 2.
CC PROSITE; PS50022; FA58C_3; 2.
CC PROSITE; PS00079; MULTICOPPER_OXIDASE1; 3.
CC Blood coagulation; Repeat; Plasma; Acute phase; Calcium;
KW

```

```

KW Signal; Glycoprotein; Sulfation.
FT SIGNAL 1 19
FT CEAIN 20 2319 POTENTIAL.
FT DOMAIN 20 349 COAGULATION FACTOR VIII.
FT DOMAIN 20 349 F5/8 TYPE A 1.
FT DOMAIN 207 349 PLASTOCYANIN-LIKE 1.
FT DOMAIN 399 730 PLASTOCYANIN-LIKE 2.
FT DOMAIN 399 730 F5/8 TYPE A 2.
FT DOMAIN 583 730 PLASTOCYANIN-LIKE 3.
FT DOMAIN 583 730 PLASTOCYANIN-LIKE 4.
FT DOMAIN 760 1640 B.
FT DOMAIN 1883 2008 F5/8 TYPE A 3.
FT DOMAIN 1883 1845 PLASTOCYANIN-LIKE 5.
FT DOMAIN 1885 2008 PLASTOCYANIN-LIKE 6.
FT DOMAIN 2008 2156 F5/8 TYPE C 1.
FT DOMAIN 2161 2313 F5/8 TYPE C 2.
FT SITE 391 392 CLEAVAGE (BY THROMBIN) (BY SIMILARITY).
FT SITE 759 760 CLEAVAGE (BY THROMBIN) (BY SIMILARITY).
FT SITE 1678 1579 CLEAVAGE (BY THROMBIN) (BY SIMILARITY).
FT SITE 1324 1325 CLEAVAGE (ACTIVATION) (BY SIMILARITY).
FT SITE 1640 1541 CLEAVAGE (ACTIVATION) (BY SIMILARITY).
FT MOD_RES 367 367 SULFATION (BY SIMILARITY).
FT MOD_RES 737 737 SULFATION (BY SIMILARITY).
FT MOD_RES 738 738 SULFATION (BY SIMILARITY).
FT MOD_RES 742 742 SULFATION (BY SIMILARITY).
FT MOD_RES 1669 1669 SULFATION (REQUIRED FOR VWF BINDING)
FT (BY SIMILARITY).
FT MOD_RES 1687 1687 SULFATION (BY SIMILARITY).
FT DISULFID 173 199 PROBABLE.
FT DISULFID 547 573 PROBABLE.
FT DISULFID 1819 1845 BY SIMILARITY.
FT DISULFID 2008 2156 BY SIMILARITY.
FT DISULFID 2161 2313 BY SIMILARITY.
FT CARBOHYD 61 61 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 233 233 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 259 259 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 423 423 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 601 601 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 880 880 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 958 958 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1015 1015 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1022 1022 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1026 1026 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1044 1044 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1076 1076 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1087 1087 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1136 1136 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1161 1161 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1192 1192 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1255 1255 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1268 1268 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1273 1273 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1274 1274 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1302 1302 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1316 1316 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1340 1340 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1378 1378 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1797 1797 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 2105 2105 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 2319 AA; 266148 MW; PD054DE051DB2A01 CRC64;

```

```

Query Match 71.1%; Score 8827; DB 1; Length 2319;
Best Local Similarity 72.5%; Pred. No. 0;
Matches 1712; Conservative 232; Mismatches 364; Indels 54; Gaps 22;

Qy 1 MQIELSTGFFCLRLPFCPSATRRYVILGAVELSWDYMQSD-LGELPVDARFPFRVPKSPFF 59
Db 1 MQIALFACFFLSLNFECSSATRRYVILGAVELSWNIQSDLLSVLHSDRFLPRMSTSPFF 60
Qy 60 NTSVYKKTLFVEFTDHLFNIAKPRPPMMGLGPTIQAEVVDVTVITLKNMASHPVSLHA 119
Db 61 NTSIMYKKTVVEYKQDLFNIAKPRPPMMGLGPTIWTVEVHTVITLKNMASHPVSLHA 120
Qy 120 VGVSVKWKASGAEYDDQTSQREKEDDKVFPGGSHYVWVLKENGPMASDPLCLTYSILS 179

```

Db 121 VGVSYKASGEDEVQDSQMEKEDDVFPBESHYYVWVLKNGPMASDPCLUTYSYXS 180
Qy 180 HVDLVKDLNSGLIGALLVCRBGSJAKEXTOLHFKPILLFAVDFDECKSWHSETKNSLMODR 239
Db 181 HVDLVKDLNSGLIGALLVCKEGLSKERTOMLYQFVLLFAVDFDECKSWHSETKNSLMODR 240
Qy 240 DAASARAWPKWHTVNGYVNRSLPGLIGCHRSYKWHVIGMGTTPVHVSIFLEGHTFLVN 299
Db 241 DSASARDWPKWHTVNGYVNRSLPGLIGCHRSYKWHVIGMGTTPVHVSIFLEGHTFLVN 300
Qy 300 HQASLESIPITFLTAQTLMDLGOFLSCHISSHCHDMEAYKVDSCPEEPQURMK-N 358
Db 301 HQASLESIPITFLTAQTLMDLGOFLSCHISSHCHDMEAYKVDSCPEEPQURMK-N 360
Qy 359 NEBAEDYODDUTDSMVVRDDNSPFIQIRSVAKHPTKWTWYIIAABEDMDYAPLV 418
Db 361 NEEMEDYODDUTDSMVVRDDNSPFIQIRSVAKHPTKWTWYIIAABEDMDYAPLV 418
Qy 419 LAPDDRSYKSVLNGPQIRGKYKVPMAVDTETREAIQESGILGPLLYGEVD 478
Db 419 PDSNGSYKSVLNGPQIRGKYKVPMAVDTETREAIQESGILGPLLYGEVD 478
Qy 479 TLLIIFKQASRPYNIYPHGITDVAPLSRRPFGVHKLKDPPIPLGPIFKYKWTIVVED 538
Db 479 TLLIIFKQASRPYNIYPHGITDVAPLSRRPFGVHKLKDPPIPLGPIFKYKWTIVVED 538
Qy 539 GPTKSDPCLTRYSSVFNMRDLASGLIGPLIICYKSVQORNGINSDKXNVLPSVF 598
Db 539 GPTKSDPCLTRYSSVFNMRDLASGLIGPLIICYKSVQORNGINSDKXNVLPSVF 598
Qy 599 DENRSWYITENIQRPLPAGVQLRDRPFOASNMHSINGVYFDSLOLVCLHEVAYWYI 658
Db 599 DENRSWYITENIQRPLPAGVQLRDRPFOASNMHSINGVYFDSLOLVCLHEVAYWYI 658
Qy 659 LSIGAQTDFLSYFFSGYTFKHKWYEDTLTLPFFSGETVFKSMENPGLMWLGCHNSDFRN 718
Db 659 LSIGAQTDFLSYFFSGYTFKHKWYEDTLTLPFFSGETVFKSMENPGLMWLGCHNSDFRN 718
Qy 719 RGMTHALLKVSCKDNTGYEDSYEDISAYLLSKNNAIDPRSPONSHPSTROQFNAT 778
Db 719 RGMTHALLKVSCKDNTGYEDSYEDISAYLLSKNNAIDPRSPONSHPSTROQFNAT 778
Qy 779 TIPENDIEKDFWFAHRTMPMKIQNVSSDMLMLRQS-PFPHGLSLSDLOBAKYETPSD 837
Db 779 TIPENDIEKDFWFAHRTMPMKIQNVSSDMLMLRQS-PFPHGLSLSDLOBAKYETPSD 837
Qy 838 DHPSCAIDNSNLSRMTFRPQLHSGDMVFTFESGLQRLNEKLGTTAATLKKLDPKV 897
Db 838 DHPSCAIDNSNLSRMTFRPQLHSGDMVFTFESGLQRLNEKLGTTAATLKKLDPKV 897
Qy 898 SSTSNNLI-STIPSNLAAGTNTSSLGPPSPVHYDSOLDTTLFGKSSPLTESGGPLS 956
Db 898 SSTSNNLI-STIPSNLAAGTNTSSLGPPSPVHYDSOLDTTLFGKSSPLTESGGPLS 956
Qy 957 LSEENDSKLPSGLMNSOESWGNVSVSTESGRLPFGKRAHCPALLTKDNALFKVSI-SL 1016
Db 957 LSEENDSKLPSGLMNSOESWGNVSVSTESGRLPFGKRAHCPALLTKDNALFKVSI-SL 1016
Qy 1017 LKTNKTSNNSA-NRKHIDGPELLIENSPPVQW-N-ILESDTEFKKVTPLIHRMLMXNA 1075
Db 1017 LKTNKTSNNSA-NRKHIDGPELLIENSPPVQW-N-ILESDTEFKKVTPLIHRMLMXNA 1075
Qy 1076 TALRLNHNKNTSSKNMEMVQKKEGPIPPDAQPDMSFFKMLTLPESARWIOPTHGN 1135
Db 1076 TALRLNHNKNTSSKNMEMVQKKEGPIPPDAQPDMSFFKMLTLPESARWIOPTHGN 1135
Qy 1136 SLNSGQSPKQVLSLGPESVQGNFSEKKNVYVKGKEFTKQVGLKEMWPPSSRNFL 1195
Db 1136 SLNSGQSPKQVLSLGPESVQGNFSEKKNVYVKGKEFTKQVGLKEMWPPSSRNFL 1195
Qy 1196 TNLKDLHNHNTNNOEKQIEBIEKKTLLIQENWVLPQIHVTGTGKNFMKQLFLLSRQW 1255
Db 1196 TNLKDLHNHNTNNOEKQIEBIEKKTLLIQENWVLPQIHVTGTGKNFMKQLFLLSRQW 1255

Db 1198 TTLSNVHENGHRNQEKNIQEBIE-KEALIBKVVLPQVHEATGSKFLKDLILLCTRQNI 1256
Qy 1256 EGSVDGAYAVLPQDFRSLNDSTNTRKHTAHFSK--KGEBENLEGLNQHTQIYKVEYACT 1313
Db 1257 --SYEVEVPVLOITISNNSTNTVQHMEHFFFKRDKETNSGLVNKTEBMYKNY--- 1311
Qy 1314 TRISPNISQOQNFVORSKRALQKQRLPLEETELEKRIIVDDTTSQWSKMKHLPSTLTQ 1373
Db 1312 -----FSQKNITQSRKALQOFL-----STQWLKTNCSQCIKQ 1349
Qy 1374 IDYNEKEGAIQTQSLDCLTRSHSIPQANRSPPIAKVGSFPPSIRPIYLRVLPQDNSS 1433
Db 1350 IDHSEKMKZF-TKSLSDS-SVIKSTTQTNSSDHIKVTSAFP---PIDLKRSPFQNKFS 1405
Qy 1434 HLPAAZY-----RKXDSGVQSSSHFLOGAKKNLSIAULTLEMTGDQREVGSIGTSATNSV 1499
Db 1406 HVQASSYIYDPKTSRIQESNNFLKTKINNPISAILPWNMFIDQGRFTSPGKSNNSV 1455
Qy 1490 TYKKVENTVLPKDPKPTSGKVLLPKVHIYQKDLFPETETNSGSPGHLDLVEGSLLOQTE 1549
Db 1466 TYKKENIIPKPTLPRESGIELLPQVSIQEBILPTETSHSGSPGHLNLAKEVFLQIKI 1525
Qy 1550 GAIKWNENRPGKVPFLRVATESAKTPSKULLDPLANDNHYGTQIPKEWKSQKSPBKT 1609
Db 1526 OPTKWKAKRHE--SIKGTESKNTSRKLLNEHAWDYHAAQIPKDMWKSQKSPBKT 1583
Qy 1610 AFKKOTILSLNACSNHATAINEGQNKPELEVWAKQGETRLCSQNPVLAKEHOREI 1669
Db 1584 SIKOEDTILSRPHGNSHISGA-NEKQWNPQRETTWVKQGTQRTCSQIPVLRKHQBEL 1642
Qy 1670 TRTTLQSDOEIDYDDTISVEMKXEDFDIYDEDENQSPRSFKKTRHYFIAAVERLWDY 1729
Db 1643 --SAFQSOEATDYDDAITIE-TIEDFDIYSEDIKQGRSPQKTRHYFIAAVERLWDY 1699
Qy 1730 MSSSEHVLRNQAQSGSVQFKKVPQBPEDTQSGTQPLRGHNEHLGLLGYIIRAEVEDN 1789
Db 1700 MSTL-HVLRNRYQDNVPOFKKVVQBPEDTQSGTQPLRGHNEHLGLLGYIIRAEVEDN 1758
Qy 1790 IMVTRQASRPYFYSLSIYEBDQROGAPRKNVFPNETKTYFKVQOHMAPTKDEF 1849
Db 1759 IMVTRQASRPYFYSLSIYEBDQROGAPRKNVFPNETKTYFKVQOHMAPTKDEF 1817
Qy 1850 DCKAWAYESVDVLEKDVHSGSLIGPLLVCHTNTLPAHGRQVTVQBFALFFTFIDETKSWY 1909
Db 1818 DCKAWAYESVDVLEKDVHSGSLIGPLLVCHTNTLPAHGRQVTVQBFALFFTFIDETKSWY 1877
Qy 1910 FTENMERNCRAPCNIOKEDPTFKENYRPHANINGIMDTLPGLVMAQDORIRWYLLSMGNN 1969
Db 1878 FTENMERNCRAPCNIOKEDPTFKENYRPHANINGIMDTLPGLVMAQDORIRWYLLSMGNN 1937
Qy 1970 ENIRSIHESGHVFTVRKKEBKALYNLYPGVFTVEMLPKAGIWRVECLIGEHLHAGM 2029
Db 1938 ENIRSIHESGHVFTVRKKEBKALYNLYPGVFTVEMLPKAGIWRVECLIGEHLHAGM 1997
Qy 2030 STLFLVYGNKQCTPLGMASGHIRDQITASGOYQWAPKARLARLHYSGSINAWSTKEPPSW 2089
Db 1998 STLFLVYGNKQCTPLGMASGHIRDQITASGOYQWAPKARLARLHYSGSINAWSTKEPPSW 2057
Qy 2090 IKVDLLAPMIINGIKTOGAROKFSSLYISQFIIMVSLDGKQWTVRGNSTGLVMPFGNV 2149
Db 2058 IKVDLLAPMIINGIKTOGAROKFSSLYISQFIIMVSLDGKQWTVRGNSTGLVMPFGNV 2117
Qy 2150 DSSGKHNIFNPPIARIYRILHPTHYRSTRLRMLMGLNCSMPLGMSKASISDAQI 2209
Db 2118 DSSGKHNIFNPPIARIYRILHPTHYRSTRLRMLMGLNCSMPLGMSKASISDAQI 2177
Qy 2210 TASSYFTNMFAWSPSKARLHQGRSNARWPCVNNPKBWLQVDFOKTMMKVTGVTTOGVKS 2269
Db 2178 TASSYFTNMFAWSPSKARLHQGRSNARWPCVNNPKBWLQVDFOKTMMKVTGVTTOGVKS 2237
Qy 2270 LLTSMYKZEFLLISSQDQHOWTLFPQNGKVKVQGNQDSFTFPVNSLDPEPLTRILRHP 2329
Db 2238 LLTSMYKZEFLLISSQDQHOWTLFPQNGKVKVQGNQDSFTFPVNSLDPEPLTRILRHP 2297

Qy	2330	QSWVHQIALRNEVLGCEAQQDY	2351	FT	DOMAIN	20	199	PLASTOCYANIN-LIKE 1.
Db	2238	QWEHQIALRLEILGCEAQQY	2319	FT	DOMAIN	207	357	PLASTOCYANIN-LIKE 2.
				FT	DOMAIN	399	730	F5/8 TYPE A 2.
				FT	DOMAIN	399	573	PLASTOCYANIN-LIKE 3.
				FT	DOMAIN	583	730	PLASTOCYANIN-LIKE 4.
				FT	DOMAIN	760	1599	B.
				FT	DOMAIN	1495	1822	F5/8 TYPE A 3.
				FT	DOMAIN	1495	1659	PLASTOCYANIN-LIKE 5.
				FT	DOMAIN	1669	1822	PLASTOCYANIN-LIKE 6.
				FT	DOMAIN	1822	1970	F5/8 TYPE C 1.
				FT	DOMAIN	1975	2127	F5/8 TYPE C 2.
				FT	SITE	391	392	CLEAVAGE (BY THROMBIN) (BY SIMILARITY).
				FT	SITE	759	760	CLEAVAGE (BY THROMBIN) (BY SIMILARITY).
				FT	SITE	1449	1450	CLEAVAGE (ACTIVATION) (BY SIMILARITY).
				FT	SITE	1490	1491	CLEAVAGE (BY THROMBIN) (BY SIMILARITY).
				FT	SITE	737	737	SULFATION (BY SIMILARITY).
				FT	MOD_RES	738	738	SULFATION (BY SIMILARITY).
				FT	MOD_RES	742	742	PROBABLE.
				FT	DISULFID	173	199	PROBABLE.
				FT	DISULFID	547	573	PROBABLE.
				FT	DISULFID	1633	1659	PROBABLE.
				FT	DISULFID	1822	1970	BY SIMILARITY.
				FT	DISULFID	1975	2127	BY SIMILARITY.
				FT	CARBOHYD	233	233	N-LINKED (GLCNAC. .) (POTENTIAL).
				FT	CARBOHYD	259	259	N-LINKED (GLCNAC. .) (POTENTIAL).
				FT	CARBOHYD	601	601	N-LINKED (GLCNAC. .) (POTENTIAL).
				FT	CARBOHYD	929	929	N-LINKED (GLCNAC. .) (POTENTIAL).
				FT	CARBOHYD	985	985	N-LINKED (GLCNAC. .) (POTENTIAL).
				FT	CARBOHYD	1025	1025	N-LINKED (GLCNAC. .) (POTENTIAL).
				FT	CARBOHYD	1111	1111	N-LINKED (GLCNAC. .) (POTENTIAL).
				FT	CARBOHYD	1181	1181	N-LINKED (GLCNAC. .) (POTENTIAL).
				FT	CARBOHYD	1208	1208	N-LINKED (GLCNAC. .) (POTENTIAL).
				FT	CARBOHYD	1245	1245	N-LINKED (GLCNAC. .) (POTENTIAL).
				FT	CARBOHYD	1265	1265	N-LINKED (GLCNAC. .) (POTENTIAL).
				FT	CARBOHYD	1335	1335	N-LINKED (GLCNAC. .) (POTENTIAL).
				FT	CARBOHYD	1408	1408	N-LINKED (GLCNAC. .) (POTENTIAL).
				FT	CARBOHYD	1611	1611	N-LINKED (GLCNAC. .) (POTENTIAL).
				FT	CARBOHYD	1919	1919	N-LINKED (GLCNAC. .) (POTENTIAL).
				FT	CONFLICT	713	713	N -> M (IN REF. 2).
				FT	CONFLICT	734	734	I -> T (IN REF. 2).
				FT	CONFLICT	792	792	G -> Q (IN REF. 2).
				FT	CONFLICT	1133	1133	E -> F (IN REF. 2).
				FT	CONFLICT	1191	1191	I -> L (IN REF. 2).
				FT	CONFLICT	1209	1209	R -> F (IN REF. 2).
				FT	CONFLICT	1437	1437	C -> G (IN REF. 2).
				FT	CONFLICT	1456	1456	F -> R (IN REF. 2).
				FT	CONFLICT	1539	1539	F -> R (IN REF. 2).
				FT	CONFLICT	1546	1546	Q -> N (IN REF. 2).
				SQ	SEQUENCE	2133	AA; 239304	MM; 152BBAS997F570DA CRC64;
								Query Match 65.0%; Score 8068; DB 1; Length 2133;
								Best Local Similarity 67.0%; Pred. No. 0;
								Matches 1577; Conservative 208; Mismatches 344; Indels 226; Gaps 17;
Qy	1	MOELSTCFCLLPFCPSATRRYVLGAVELSWDMQSD-IGELPDARFPRVPKSPFF	59	Qy	1	MOELSTCFCLLPFCPSATRRYVLGAVELSWDMQSD-IGELPDARFPRVPKSPFF	59	
Db	1	MOELSTCFCLLPFCPSATRRYVLGAVELSWDMQSD-IGELPDARFPRVPKSPFF	60	Db	1	MOELSTCFCLLPFCPSATRRYVLGAVELSWDMQSD-IGELPDARFPRVPKSPFF	60	
Qy	60	NTSVVYKTLVEFTDHLFNIAKPPPMWGLLGPITQAEVVDVTVTLKNWASHPVSLHA	119	Qy	60	NTSVVYKTLVEFTDHLFNIAKPPPMWGLLGPITQAEVVDVTVTLKNWASHPVSLHA	119	
Db	61	GESVLYKTVFVEFTDHLFNIAKPPPMWGLLGPITQAEVVDVTVTLKNWASHPVSLHA	120	Db	61	GESVLYKTVFVEFTDHLFNIAKPPPMWGLLGPITQAEVVDVTVTLKNWASHPVSLHA	120	
Qy	120	GVSVYWKASGEAEYDDQTSQREKEDKVPFGSGSHYVWVLKENGPMASDPLCTYSYLS	179	Qy	120	GVSVYWKASGEAEYDDQTSQREKEDKVPFGSGSHYVWVLKENGPMASDPLCTYSYLS	179	
Db	121	GVSVYWKASGEAEYDDQTSQREKEDKVPFGSGSHYVWVLKENGPMASDPLCTYSYLS	180	Db	121	GVSVYWKASGEAEYDDQTSQREKEDKVPFGSGSHYVWVLKENGPMASDPLCTYSYLS	180	
Qy	180	HYDLVKDLNSGLIGALLVCRGSLAKGKTQTLHKILLFAVDFDEGSKSHSETKNSLMQDR	239	Qy	180	HYDLVKDLNSGLIGALLVCRGSLAKGKTQTLHKILLFAVDFDEGSKSHSETKNSLMQDR	239	
Db	181	HYDLVKDLNSGLIGALLVCRGSLAKGKTQTLHKILLFAVDFDEGSKSHSETKNSLMQDR	240	Db	181	HYDLVKDLNSGLIGALLVCRGSLAKGKTQTLHKILLFAVDFDEGSKSHSETKNSLMQDR	240	
Qy	240	DAASARAFKMTVNGYVNRSLPGLIGCRKSVYWHVIGMCTTPEVHSTLEBHTFLVFN	299	Qy	240	DAASARAFKMTVNGYVNRSLPGLIGCRKSVYWHVIGMCTTPEVHSTLEBHTFLVFN	299	
Db	241	DPAPARAFKMTVNGYVNRSLPGLIGCRKSVYWHVIGMCTTPEVHSTLEBHTFLVFN	300	Db	241	DPAPARAFKMTVNGYVNRSLPGLIGCRKSVYWHVIGMCTTPEVHSTLEBHTFLVFN	300	

QY 300 HQASLEISPIELTQATLMDLQPLLSCHISSHOEDGMBAYVVKVDSQPBEPQLEMKN 359
DB 301 HQASLEISPLTQATLMDLQPLLSCHSSHGGMEAVRVESCAEBPQURKAD 360
QY 360 REAEDYDDDLTSEMDVDFDDNSPFIQIRSAVAKKHPTWVHYIAABEEDWDYAPLV 419
DB 361 EE-EDYDDNLYSDMDVRLDGDVSPFFIQRSAVAKKHPTWVHYISABEEDWDYAPV 419
QY 420 APDRSYKQYLANGPQIRGRKYKRYFWAYTDETFTATREAIQHEGILGLPLLYGBVGD 479
DB 420 SPDSRYSKSIYLSGFCQIRGRKYKRYFAYTDVTFKRAKAIPIYESGILGLPLLYGBVGD 479
QY 480 LLIIFKQASRPNIYPHGTIDVRPLSRPLKGVGHLKDFPLPBGRIKPKYKWTVVVEOG 539
DB 480 LLIIFKQASRPNIYPHGTIDVSALHPGLRGKGVGHLKDXPLPGETTFYKRWTVVVEOG 539
QY 540 PTKSDPCLTRYVSSFMNERDIASGLIGPLLICYKESVDQRGNQIMSKRNVLFSVFD 599
DB 540 PTKSDPCLTRYVSSSINLEKDLASGLIGPLLICYKESVDQRGNQIMSKRNVLFSVFD 599
QY 600 ENRSWYLTENIQRLPNAGVOLEDPPFOASINMHISINGVFDLSQLSVCLHEVAWYIL 659
DB 600 ENQSWYLAENIQRLPNAGVOLEDPPFOASINMHISINGVFDLSQLSVCLHEVAWYIL 659
QY 660 SIGAOTDFLVSFGYTFKHVYEDTLTFPFSGETVMSMENPGLWILGCHNSDPNR 719
DB 660 SVGAOTDFLVSFGYTFKHVYEDTLTFPFSGETVMSMENPGLWILGCHNSDLNR 719
QY 720 GMTALLKVSCKNTGDIYEDSYEDI SAYLISKNAIEPRSFSONSRHPSSTRQKFNATT 779
DB 720 GMTALLKVSCKRDIYDNTYEDI PCFLLSGKNVIEPRSFSONSRHPSSTRQKFNATT 779
QY 780 IPENDIKTPFAHRTPMKAIQNVSSDMLMLRQSPFHGLSLSLDQAKYETPDDP 839
DB 780 SPEDDVE-LDPQSGERTQALELSVPGDGMGLLQNPAPHGSSSDLEQEARNE--ADDY 836
QY 840 SPGAIDSNISISEMTHRPQLHSGDMVFTPESGQLRLNEKJCTTATELKLKDFKVS 899
DB 837 LFGARENTAPSAARURPELHSAERVLTEP-----EK-----ELKLDKMS 882
QY 900 TSNNLIS--TIPSDNLAAGTNTSLGPPSPVHYDSQDITFLFKKSSPFTESGGPLSL 957
DB 883 SSDLLKCSFTIPSDLSAETERTSLGPPHPQVNFPSQIGAIVLGKXSSHFGAGVPLGS 942
QY 958 SHENNDKLESGLMNSQESGAGKVSSTESGRFLPKXRAHGPALLTKONALFKVUSILL 1017
DB 943 TEEDH-----ESSLGENVSPVESDGIPEKERAHGPASLTAKDDVLFKVNISLV 989
QY 1018 KTNKTSNNSATNRKTHIDGPSLILINSPSVWQNLIESDTEFKKVTPLIHDRMLDKNATA 1077
DB 990 KTKARVYLKTKRHIDDAALLTENRASA-----TFMDKNVTA 1028
QY 1078 LRLNHSNKTTSKXNMVQOKKEGPIPDQONPMSFFKMLFUPESARWQRTHGKNSL 1137
DB 1029 SGLNHSN-----WIKGPIKGNPL 1047
QY 1138 NSGQGPSKQIVSLGPEKSVBQNFLEKKNVVKGBFTDVGLEKVMVPPSSRNLFJTN 1197
DB 1048 SSREGSPPELLTSSGSGSVKSGQSGQRIRVAVEEELSUG---KENMLPNSLTFJTN 1104
QY 1198 LDNLHENHNTQEKKIQREIEKETLIQNVVLPQIHTVVTGTFNPMKNLFLLSRQNVG 1257
DB 1105 SADVQGNTHSGOKKSREEMEREXLVQEKVDLPQVYIATGFKNLFNIFHQSTEPSVEG 1164
QY 1258 SYDGAYAPVLODFRSLNSTNTKXHTAHPSKKGBEENLEGNGNOTKQIVEKYACTTRIS 1317
DB 1165 FDCGSHAPVPODSRGLNSAERAETHIAHFAAREEAPLEAPCNR-----1210
QY 1318 PNTSQNFVQSKKAIQKFRPLRETELEKLIIVDDTSTQWSKNMKHLTPSLTLQIDYN 1377
DB 1211 -QFGFSAPVRVKQSLKQIRLEFLEBEIKPBGVNLNATSTRWS-----1252

RESULT 4

PAS_HUMAN

ID PAS_HUMAN

AC P12259; Q14285;

STANDARD;

PRT: 2224 AA.

QY 1378 EXEKGAITQSPLSDCILTRSHSIPQANRSPLIATAKVSFSPISRIPIYLTRVLFDQNSHLPA 1437
DB 1253 -----1252
QY 1438 ASYRKDSQVQSSSHFLQAKAKNNLSLAILTLEMTGDQREVSGLSITSATNSVYKVENT 1497
DB 1253 -----ESSPIQAGAKNNLSLPTLITLEMAGGQGIKISALGSAAGPLASGKLEKA 1301
QY 1498 VLPKPDLPKTSKGVKLLPKVHIYOKDLRPTETSNQSPGHLDLVEGSLLOCTEGAIKNFA 1557
DB 1302 VSSAGLSASAKABFLPKVRVEREDLLPQKISNVSCAHGDLQOEIQLQKTRGPVNLKV 1361
QY 1558 NRPKGVPFLRVATESAKTFSKLLDPLAWDNHYGTQIPKEEMKQSKSPEKTAFKKOTI 1617
DB 1362 NRPK-----RTPSKLLGP-----PMPK-EWESLEKSPKSTALRTKDII 1398
QY 1618 -LSLNACENHAIANBEGONKPEIIVTWAKQGRTERLCSOMPVLKRHOREITRRLQS 1676
DB 1399 SLPLDSESHNSIAAKNEGCAETQREAAWTKQGGPGRLCAPKPPVILRRHQDLSLPTFQ 1458
QY 1677 DOEEIDYDDTISVEMKKKDFDIYDEDEENQSPRSFOKKTREHYPIAAVERLWDYGMSSPHV 1736
DB 1459 EEDKMDYDDIFSTETKGEDFDIYGEDENQDPRSFQKTRHYFIAAVEQLWDYGMSSPRA 1518
QY 1737 LRNRAQSGSVQPKKVVFOEFTDGSFTQPLRYGELNEHGLLGPYIRAEVEDNIMVTFRN 1796
DB 1519 LRNRAQNGEVRPKKVVREFAFGSFTQPSYRGELNKHGLLGPYIRAEVEDNIMVTFKN 1578
QY 1797 QASRPVSFYSLLISYEDQRCGAEPKRVKNETKTYFWKVOHNAPTKDEDFCKAWAY 1856
DB 1579 QASRPYSFYSSLLISYDDQCGAEPRHNFVQNETRTYFWKVOHNAPTDEDFCKAWAY 1838
QY 1857 FSDVLEKDVHSGLIGPLLCHTNTLNPAGHGRQVTVQEPFALFFITIDETKSWYFTENMR 1916
DB 1639 FSDVLEKDVHSGLIGPLLICRANTLNAAGHGRQVTVQEPFALFFITIDETKSWYFTENMR 1698
QY 1917 NCRAPCNIQMEDTFFKENYRPHAINGVIMDTPLGLVNAQQRIRWYLLSGNSENIHSIH 1976
DB 1699 NCRAPCHLQMEDPTLKENYRPHAINGVIMDTPLGLVNAQQRIRWYLLSGNSENIHSIH 1758
QY 1977 FSGHVFTVRKEEYKMALYNLYSGVETVEMLPKSKAGIWEVECLIGPHLHAGMSTFLVY 2036
DB 1759 FSGHVFTVRKEEYKMAVYNLYPGVFTVEMLPKSKAGIWEVECLIGPHLHAGMSTFLVY 1818
QY 2037 SNKQOTPLGNASGHIRDFOITASQYQWAPKLARLHYSGSINAWSTKPPSWIKVDLLA 2096
DB 1819 SKEQAPLGNASGHIRDFQITASQYQWAPKLARLHYSGSINAWSTKDPHSHWIKVDLLA 1878
QY 2097 PMIHGKTQCAROKFSSLYISQFIIMYSLDGKKWQYRGNSTGTLMVFFGNVDSSGKIH 2156
DB 1879 PMIHGINTQCAROKFSSLYISQFIIMYSLDGRNMQSYRGNSTGTLMVFFGNVDASGKIH 1938
QY 2157 NIENPPIIARVIRLHPHYSIRSTRVLELMGCDLNSCSMPLGMSKASISDAQITASSYFT 2216
DB 1939 NIFPPIVARVIRLHPHYSIRSTRVLELMGCDLNSCSMPLGMSKASISDAQITASSHLS 1998
QY 2217 NMFATWSPSKARLHLOGRSNAWRPQVNNPKWLQVDFQKTMKVTVGTVTQGVKSLTSMYV 2276
DB 1999 NIFATWSPSQARLHLOGRTNARPRVSAEELQVDI-QKTVKVTGITTTQGVKSLTSMYV 2058
QY 2277 KEFLISSQDQHWTLRFQNGKVKVFCQNGQDSFTFVNVNSLDPPLLTRYLRIHPQSVWHQI 2336
DB 2059 KEFTLVSSQDGRWTLFLQDGHFTKVFQNGQDSSTFVNVNLDPLPLFTRYLRIHPTSWAQHI 2118
QY 2337 ALRVEVLGCEAQLDY 2351
DB 2119 ALRVEVLGCEAQLDY 2133

DT 01-OCT-1989 (Rel. 22, Created)
 DT 01-FEB-1994 (Rel. 28, last sequence update)
 DT 15-MAR-2004 (Rel. 43, last annotation update)
 DE Coagulation factor V precursor (Activated protein C cofactor).
 GN F5.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 [1]
 RN SEQUENCE FROM N.A., AND VARIANT MET-1764.
 RP MEDLINE=92232658; PubMed=1567832;
 RX Cripe L.D., Moore K.D., Kane W.H.;
 RA "Structure of the gene for human coagulation factor V";
 RL Biochemistry 31:3777-3785(1992).
 [2]
 RN SEQUENCE FROM N.A., AND VARIANT GLU-925.
 RP MEDLINE=87260886; PubMed=3110773;
 RX Jerry R.J., Pittman D.D., Toole J.J., Kriz R.W., Aldape R.A.,
 RA Hewick R.W., Kaufman R.J., Mann K.G.;
 RT "Complete cDNA and derived amino acid sequence of human factor V";
 RL Proc. Natl. Acad. Sci. U.S.A. 84:4846-4850(1987).
 [3]
 RN SEQUENCE OF 1-1600 FROM N.A., AND VARIANTS GLU-925 AND ILE-1285.
 RP MEDLINE=88107560; PubMed=2827731;
 RX Kane W.H., Ichinose A., Hagen F.S., Davie E.W.;
 RA "Cloning of cDNAs coding for the heavy chain region and connecting
 RT region of human factor V, a blood coagulation factor with four types
 RT of internal repeats";
 RL Biochemistry 26:6508-6514(1987).
 [4]
 RN SEQUENCE OF 1188-1215 AND 1315-2224 FROM N.A.
 RX MEDLINE=86313663; PubMed=3092220;
 RA Kane W.H., Davie E.W.;
 RT "Cloning of a cDNA coding for human factor V, a blood coagulation
 RT factor homologous to factor VIII and ceruloplasmin";
 RL Proc. Natl. Acad. Sci. U.S.A. 83:6800-6804(1986).
 [5]
 RN PARTIAL SEQUENCE FROM N.A.
 RP TISSUE=Fibroblast;
 RX MEDLINE=93203619; PubMed=8454869;
 RA Shen N.L.L., Fan S.-T., Pyati J., Graff R., Lapolla R.J.,
 RA Edgington T.S.;
 RT "The serine protease cofactor factor V is synthesized by
 RT lymphocytes";
 RL J. Immunol. 150:2992-3001(1993).
 [6]
 RN SULFATION.
 RP MEDLINE=94264012; PubMed=8204629;
 RX Pittman D.D., Tomkinson K.N., Michnick D., Seligsohn U.,
 RA Kaufman R.J.;
 RT "Posttranslational sulfation of factor V is required for efficient
 RT thrombin cleavage and activation and for full procoagulant activity";
 RL Biochemistry 33:6952-6959(1994).
 [7]
 RN SULFATION.
 RP MEDLINE=90366699; PubMed=2168225;
 RX Hortic G.L.;
 RA "Sulfation of tyrosine residues in coagulation factor V";
 RL Blood 76:945-952(1990).
 [8]
 RN X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS) OF 2065-2224.
 RP MEDLINE=20052169; PubMed=10586886;
 RX Macedo-Ribeiro S., Bode W., Huber R., Quinn-Allen M.A., Kim S.W.,
 RA Ortel T.L., Bourenkov G.P., Bartunik H.D., Stubbs M.T., Kane W.H.,
 RA Fuentes-Prior P.;
 RT "Crystal structures of the membrane-binding C2 domain of human
 RT coagulation factor V";
 RL Nature 402:434-439(1999).
 [9]
 RN VARIANT MET-1764.
 RP MEDLINE=95179146; PubMed=7874144;
 RX Bayston T.A., Ireland H., Olds R.J., Thein S.L., Lane D.A.;
 RA

RT "A polymorphism in the human coagulation factor V gene";
 RL Hum. Mol. Genet. 3:2085-2085(1994).
 [10]
 RN VARIANT APCR GLN-534.
 RP MEDLINE=94217810; PubMed=8164741;
 RX Bertina R.M., Koelman B.P.C., Koster T., Rosendaal F.R.,
 RA Dirven R.J., de Ronde H., van der Velden P.A., Reitsma P.H.;
 RT "Mutation in blood coagulation factor V associated with resistance to
 RT activated protein C";
 RL Nature 369:64-67(1994).
 [11]
 RN VARIANTS ILE-1285 AND ARG-1327.
 RP MEDLINE=96351768; PubMed=8713778;
 RX Lunghi B., Iacoviello L., Gemmati D., Dilasio M.G., Castoldi E.,
 RA Pinotti M., Castaman G., Redaelli R., Mariani G., Marchetti G.,
 RA Bernardi F.;
 RT "Detection of new polymorphic markers in the factor V gene:
 RT association with factor V levels in plasma";
 RL Thromb. Haemost. 75:45-48(1996).
 [12]
 RN VARIANT APCR GLY-334, AND VARIANT LYS-513.
 RP MEDLINE=98122763; PubMed=9454741;
 RX Chan W.P., Lee C.K., Kwong Y.D., Lam C.K., Liang R.;
 RA "A novel mutation of Arg306 of factor V gene in Hong Kong Chinese";
 RL Blood 91:1135-1139(1998).
 [13]
 RN VARIANT APCR THR-334.
 RP MEDLINE=98122764; PubMed=9454742;
 RX Williamson D., Brown K., Luddington R., Baglin C., Baglin T.;
 RA "Factor V Cambridge: a new mutation (Arg306-to-Thr) associated with
 RT resistance to activated protein C";
 RL Blood 91:1140-1144(1996).
 [14]
 RN VARIANTS HIS-107; THR-413; LYS-513; SER-809; THR-817; ARG-858;
 RP ARG-865; GLU-925; GLN-1146; ALA-1530; SER-1685; VAL-1749; MET-1764;
 RP ILE-1820 AND GLY-2222, AND VARIANT APCR GLN-534.
 RX MEDLINE=99318093; PubMed=10391209;
 RA Cargill M., Altshuler D., Ireland J., Sklar P., Ardlie K., Patil N.,
 RA Shaw N., Lane C.R., Lim E.P., Kalyanaram N., Nemesh J., Ziaugra L.,
 RA Friedland L., Rolfe A., Warrington J., Lipshutz R., Daley G.Q.,
 RA Lander E.S.;
 RT "Characterization of single-nucleotide polymorphisms in coding regions
 RT of human genes";
 RL Nat. Genet. 22:231-238(1999).
 [15]
 RN ERRATUM.
 RP Cargill M., Altshuler D., Ireland J., Sklar P., Ardlie K., Patil N.,
 RA Shaw N., Lane C.R., Lim E.P., Kalyanaram N., Nemesh J., Ziaugra L.,
 RA Friedland L., Rolfe A., Warrington J., Lipshutz R., Daley G.Q.,
 RA Lander E.S.;
 RL Nat. Genet. 23:373-373(1999).
 [16]
 RN VARIANT APCR HIS-2102.
 RP MEDLINE=21847288; PubMed=11858490;
 RX Schrijver I., Houissa-Kastally R., Jones C.D., Garcia K.C.,
 RA Zehnder J.L.;
 RT "Novel factor V C2-domain mutation (R2074H) in two families with
 RT factor V deficiency and bleeding";
 RL Thromb. Haemost. 87:294-299(2002).
 CC -!- FUNCTION: Coagulation factor V is a cofactor that participates
 CC with factor Xa to activate prothrombin to thrombin.
 CC -!- SUBUNIT: Factor Va is composed of a heavy chain and a light
 CC chain, noncovalently bound. The interaction between the two chains
 CC is calcium-dependent.
 CC -!- DOMAIN: Domain B contains 35 x 9 AA tandem repeats, and 2 x 17 AA
 CC repeats.
 CC -!- PTM: Thrombin activates factor V proteolytically to the active
 CC cofactor, factor Va (formation of a heavy chain at the N-
 CC terminus and a light chain at the C-terminus).
 CC -!- PTM: Sulfation is required for efficient thrombin cleavage and
 CC activation and for full procoagulant activity.
 CC -!- DISAAS: Defects in F5 are the cause of Owren parahemophilia
 CC [MIM:227400], an hemorhagic diastesis.

Db 389 KHYKKVYKQYQDSFTKRIENPNKNGDILGPVIRAQVRDTLKIIVFANWASRSYSIVPH 448
Qy 498 GIT-----DVRPLYSRPLKGVXHLKDPFPLPGEIKYKATVTVEDGPTKSDPRCLTRY 551
Db 449 GVTFSPVEDDYNSSSTSDNNTMIR-----AVOPGETYTYKWNILESDEPTENDAQCLTRP 503
Qy 552 YSSFNVERDIASGLIGPLLCYKESVDQRGNQIMSKRNVLFSVFDENRSWYLTENIQ 611
Db 504 YYSNVDIRDIASGLIGLLCKSRSLDKRGIQRTADIEQKAVPAVDPENKSWIEDNIY 563
Qy 612 RFLPNACVQLEDEFEQASNIWMSINGVYFBSL-QLSVCLHEVAYWILSTCAQTDLSV 670
Db 564 KFCBNPKVKDDPKFYENIMSTINGVYFBSIPTLGFCFDTQWMEFCSVRTHDNILTI 623
Qy 671 FFSGVTPKHVMYEDTTLFPFSGBTVMSENPGWLWILGCHNGDFRNRGTMALLKYSSC 730
Db 624 HFTGHSFYGKHEDTTLFPMRGESVTVMNVGTWMLTMMNSNPRKXKLQKFRDYKC 683
Qy 731 DKNTG-DYVEDSYEDISAYLL-----SKNAIAP-----RSFSQNS 765
Db 684 IRDDDEDGEIYRPSSTTLTRKWHDSSENKEBENDCEYDQDLASVLGIRSFNRS 743
Qy 766 RHPSTROQFNATIP-ENDIEKIDPWFAHRTMPKIQNVSSOLLMLLRQSPHGLSL 824
Db 744 LY--QEDDEFNLALALENNEFTFP-----STD-----RAVDSNSSP 780
Qy 825 SLOQAKVETFSDD---PSGAIDSNNS-----LSEWTRFPOLHSHGDMVFP-ESG 873
Db 791 GNISRAPANTTEPRKIILPHFEATKAGSPRHGTGLVNLVLRRTQHSDDPYSEDPIENP 840
Qy 874 IQ-----LRLNE-----KLGTAAATE--LKLDPKVSSTSNNLISIP 909
Db 841 LQSVITG-SLLPFGTEGRNREHPKHFKAGRDQAAXHRFSOMEPFAHKTGRH----IS 896
Qy 910 SDNLAACTNSSLGPPMPHYDSLODTTLPGKSSPLTESGGPLSSEHN-----962
Db 897 QDN-----SSSSSMGPLE-----DLSDDLLEK-DESTINGKWHLVSEKGSYIIVQDA 945
Qy 963 DSKILESGLMNSQES---SWGKNVSTSGRLFKGKRAFGPALLAKDNALFKVSIISLKT 1019
Db 946 DEDMAVKNLPNNFQNASWGENIFTNK---HGKRGHPFVTRH-----KLJQE 993
Qy 1020 NKTNSNS-----ATNRKTHIDGPSLLIENSPSWQNIL-ESDTEFKVTPLIHDM 1069
Db 994 RQDEGNSILKGRFFTRTRRKKERKPVHIVLSPRSFNLGEANTFPS-----1043
Qy 1070 LMDKNATALRLNEMNKITSSKNMVMVQKKGSP-PPDAQN---PDMSPFKMLPESAR 1126
Db 1044 ---DRQNHSLLHESNET-----PPFDLNOTFPMXNLSLASHPD---1082
Qy 1127 WIQRTHGKNSLN-SGQGPSPKQIVSLGPKSVEGQNFLSERKNVYVGKEFTKDVGLKEM 1185
Db 1083 ----HNQNLNDTHQTSF-----LDLYGT 1103
Qy 1186 VFPSS--RNLFTNLNDLHNTHNQEKKIQEIEKKEKTELIOENVVLPOIHVVTGKNFM 1243
Db 1104 VTFDEPYQAPIQDLDETHSTAVPSHQSSLPEPI-----QXH-----1140
Qy 1244 KXNLLILSTQNEGVSDGAVAPVLQDFRSLNDSNTRTKXHTAHS---KKGBEENLEG 1300
Db 1141 -----DYD-----LRKASPTDVSEMPFSLKLKAGHRTTSPD- 1173
Qy 1301 NQTKQIVEYACT-----TRISPNTSQQNFVQSRKALKQFRLPLEETEL-----BK 1348
Db 1174 NQTSLSPELSQTTLSDPGHWTLSPDLSQTTLSPLDLSHT-TLSP---DLGHTTLPDLSHT 1230
Qy 1349 RIIVDTSTOWSNKMKHLPST-TQIDYNEKEKAITQSPDLSCLTRSHSIPOANRSLP 1408
Db 1231 TLSPDLSQTTLPDLSHTTLLS-----PDLGHTTLPDLSHTTLPDLSHT-D 1280
Qy 1409 IAKVSSFPSTRIPIYRVLFDQDNSSHLPAASYKKKDSG-VQESSHFLQGAKKNNLSLAL 1467
Db 1281 LSHTTLSPLDGLHTTLPDLSQTTLSPLDGLHTTLPDLSQTTLSPLDGLHTTLPDLSHTTLL 1340

RESULT 6

FAS_BOVIN

ID_FAS_BOVIN STANDARD; PRT; 2211 AA.

Qy 1468 T--LEMTGDQREVSGSLCTSATNSVTYKKVENTVLP-----KFDLPKTS-----GKVEL 1513
Db 1341 SPDLSSHITLSPDLGHTTTLSPDLSQT-----TLSPDLGHMTLSPDLSSHITLSPDLGHTTTL 1394
Qy 1514 LPKVHIYOKOLFPTETENGSGPHGLDLVEGSLLOGTEGAIKWEANRPGKVPFLFVAVTSS 1573
Db 1395 SP-----DLSHT-TLSPDLGHMTLSPD---LGQTTLSLDFGQTTLSPDLSSHMTLSSELS 1444
Qy 1574 AKTSPKLL-----DPLAW-----DNHYGTQIPKEEMKSEQSPKPTAFKK 1613
Db 1445 HETLSPDLSQVTLSPDLSSEIPFSPDLMQTTLSSDLNHTTTLSPDLRQSPDPDPKTSY-- 1502
Qy 1614 KDTILSNACENSHAIAGINEGQNKPEIEVTAQKQSTERLCSQ-----NPPVLKXHQ 1666
Db 1503 -----ISESSQSVLPFPFGQTSPPDLQGRSPSPSHSTLNNFTIPRPFNPMVAVGLS 1554
Qy 1667 RE-----LTRITLQSDQES-----IDYDDDTISVENKKEDPDYDEDNQSB----- 1707
Db 1555 RDDGDYVEIIPROQFENSEEDYVKIDYVXDPPYQTDVRT-----DINSRNPDTAAW 1608
Qy 1708 --RSFOKTRHYFTAAVERLWDYCMSSSPHVLNRQAQSGVPO---PKVVVFOBFTDGSF 1762
Db 1609 YLRNNGNRNYIYIABEELSWSYK-----FTQREBIDDDVPEHTIYKVVFRKYLDSFT 1662
Qy 1763 TQPLYRGELNEHLGLGFIYRAEVEDNIMVTRNQAQSPYSPYSSLSIYE-----ED 1814
Db 1663 TKLDPGRGEYEBHLGILGFIIRAEVDVVIQVAFKNLASRPSYSLHAHGLSYEKSSGKTYED 1722
Qy 1815 ORQAAERPKNFVKNEIKTYFKVCHHMAPKDEFCKAKAYFSDVLEKDVHSLGLPL 1874
Db 1723 DSPWFCKEDNAVQNSSYTYVWHATERSGPSPGSAACRWAYIYSAVNPEDKDIHSLGLPL 1782
Qy 1875 LVCHTNTLNPAHQVTVQBPALPFTTIDETKSWYFTENMERNCRAPCNIQMEDPTFKEN 1934
Db 1793 LICRGTLKENNMPVDMREFFVLFFWVDEKKSWMYKFTESNR-----LTSSEVENS 1836
Qy 1935 YRFHAINGYIMDTLPGLVMAODQRIWYLLNGNENIHSIHFGSHVTVTRKKSEYKVAL 1994
Db 1837 HKFLTAINGMIYN-LFGLFMVQEWRVRLHLLNLGGSRDHVVVHFGQTLLENGTQQHQLG 1895
Qy 1995 YNLPYGFETVEMLPSPKAGIWRVECLIGEHLHAGMSTLFLVYSNKCOTPLGWSGHIRDF 2054
Db 1896 WPLLPGSFPTLEMTKTSKAGWMLLDTVEGENQAGNQTFFLIIDRECKPMGLSLGIADS 1955
Qy 2055 QITASQYQGWAPKLARLHYSGSINAWSTKEPFS-----WIKVDLLAPMIHGIKTQG 2107
Db 1956 QIKASEFWGHQPKLARLNNNGSYNAWIT-DKFSGESNSKZWIOVDMQREVVFQIQTOG 2014
Qy 2108 ARQKFSLSYISQFTIMYSLDGKWKQTYGNSGTGLIMVFFGNVDSGIXHNINPPIARY 2167
Db 2015 AKYLLKSYTYTTFPNVAYSSDQNRWIRFKGNSTKNMYFNNGSDASTITENOFDPPVARY 2074
Qy 2168 IRLHETHYSIRSTRVRLVLMGCDLNSCMPLGWSKAIQDAQITASSYPTNNFAT--WSPSK 2226
Db 2075 IRISTESINPALRELELOGCEVNGCCSTPLGVESNINKNEQITASSFKKSWMGD/WBPFR 2134
Qy 2227 ARHLHQGRSNARPOVNNPKQLQVDFQTKMKTGVTTCQVKSLTSMYVKEFLISSQD 2286
Db 2135 AFLNAQGRVNAWQAKANNNNQWLIQIDLKIKKITAITTOGCKSLSESMVKRYTYIQSDR 2194
Qy 2287 GHQWTLFPQNGV--KVTFQGNQDSFTPVNSLDPDLLTYLRIHPQSVWQHIALRMEVLG 2344
Db 2195 GYEMKSYREKSMWDKIPEGNNNKHGKVNFPNPIISRFIRIIPKWNQSIJALRELF 2254
Qy 2345 CEAQDIY 2351
Db 2255 C---DIY 2258

AC	Q28107; Q28108;	OS	FT	DOMAIN	30	327	F5/8 TYPE A 1.
DT	01-NOV-1997 (Rel. 35, Created)		FT	DOMAIN	30	193	PLASTOCYANIN-LIKE 1.
DT	01-NOV-1997 (Rel. 35, Last sequence update)		FT	DOMAIN	203	327	PLASTOCYANIN-LIKE 2.
DT	15-MAR-2004 (Rel. 43, Last annotation update)		FT	DOMAIN	348	686	F5/8 TYPE A 2.
DE	Coagulation factor V precursor (Acrivated protein C cofactor).		FT	DOMAIN	348	525	PLASTOCYANIN-LIKE 3.
GN	35.		FT	DOMAIN	535	686	PLASTOCYANIN-LIKE 4.
OS	Bos taurus (Bovine).		FT	DOMAIN	696	1564	B.
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		FT	DOMAIN	1124	1151	2 X 14 AA TANDEM REPEATS.
OC	Mammalia; Eutheria; Cetiartiodactyla; Ruminantia; Pecora; Bovidae;		FT	REPEAT	1124	1137	1-1.
OC	Bovidae; Bovinae; Bos.		FT	REPEAT	1138	1151	1-2.
ON	NCBI_TaxID=9913;		FT	DOMAIN	1188	1453	[AS]-L-S-P-D-[LP]-[GS]-Q-[TE].
OX	11;		FT	REPEAT	1188	1196	2-1.
RP	SEQUENCE FROM N.A.		FT	REPEAT	1197	1205	2-2.
RP	TISSUE=Liver;		FT	REPEAT	1206	1214	2-3.
RX	MEDLINE=92147639; PubMed=1737753;		FT	REPEAT	1215	1223	2-4.
RA	Guineto E.R., Esmen C.T., Mann K.G., Macgillivray R.T.;		FT	REPEAT	1224	1232	2-5.
RT	"The complete cDNA sequence of bovine coagulation factor V.;"		FT	REPEAT	1233	1241	2-6.
RL	J. Biol. Chem. 267:2971-2978(1992).		FT	REPEAT	1242	1250	2-7.
RL	12]		FT	REPEAT	1251	1259	2-8.
RP	DISULFIDE BONDS.		FT	REPEAT	1260	1268	2-9.
RP	MEDLINE=95034740; PubMed=7947716;		FT	REPEAT	1269	1277	2-10.
RA	Xue J., Kalafatis M., Silveira J.R., Kung C., Mann K.G.;		FT	REPEAT	1278	1286	2-11.
RT	"Determination of the disulfide bridges in factor Va heavy chain.;"		FT	REPEAT	1287	1295	2-12.
RL	Biochemistry 33:13109-13116(1994).		FT	REPEAT	1296	1304	2-13.
CC	-1- FUNCTION: Coagulation factor V is a cofactor that participates		FT	REPEAT	1305	1313	2-14.
CC	with factor Xa to activate prothrombin to thrombin.		FT	REPEAT	1314	1322	2-15.
CC	-1- SUBUNIT: Factor Va is composed of a heavy chain and a light		FT	REPEAT	1323	1331	2-16.
CC	chain, noncovalently bound. The interaction between the two chains		FT	REPEAT	1332	1340	2-17.
CC	is calcium-dependent.		FT	REPEAT	1341	1349	2-18.
CC	-1- DOMAIN: Domain B contains 29.5 X 9 AA tandem repeats, and 2 X 14		FT	REPEAT	1350	1358	2-19.
CC	AA repeats.		FT	REPEAT	1359	1367	2-20.
CC	-1- PMT: Thrombin activates factor V proteolytically to the active		FT	REPEAT	1368	1376	2-21.
CC	cofactor, factor Va (formation of a heavy chain at the N-		FT	REPEAT	1377	1385	2-22.
CC	terminus and a light chain at the C-terminus).		FT	REPEAT	1386	1394	2-23.
CC	-1- PMT: Sulfation is required for efficient thrombin cleavage and		FT	REPEAT	1395	1403	2-24.
CC	activation and for full procoagulant activity (By similarity).		FT	REPEAT	1404	1412	2-25.
CC	-1- SIMILARITY: Contains 3 F5/8 type A domains.		FT	REPEAT	1422	1430	2-27.
CC	-1- SIMILARITY: STRONG, TO COAGULATION FACTOR VIII.		FT	REPEAT	1431	1439	2-28.
CC	-----		FT	REPEAT	1440	1444	2-29 (PARTIAL).
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration		FT	REPEAT	1445	1453	2-30.
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation		FT	DOMAIN	1569	1890	F5/8 TYPE A 3.
CC	the European Bioinformatics Institute. There are no restrictions on its		FT	DOMAIN	1569	1738	PLASTOCYANIN-LIKE 5.
CC	use by non-profit institutions as long as its content is in no way		FT	DOMAIN	1748	1890	PLASTOCYANIN-LIKE 6.
CC	modified and this statement is not removed. Usage by and for commercial		FT	DOMAIN	1894	2048	F5/8 TYPE C 1.
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/		FT	DOMAIN	2053	2208	F5/8 TYPE C 2.
CC	or send an email to license@sb-sib.ch).		FT	SITE	741	742	CLEAVAGE (BY THROMBIN) (BY SIMILARITY).
CC	-----		FT	SITE	1034	1035	CLEAVAGE (BY THROMBIN) (BY SIMILARITY).
DR	EMBL; M81440; AAA30512.1; -;		FT	SITE	1564	1565	CLEAVAGE (BY THROMBIN).
DR	EMBL; M81441; AAA30513.1; -;		FT	DISULFID	167	193	
DR	PIR; A42580; KFPO5.		FT	DISULFID	248	329	
DR	HSP; P12259; LCZT.		FT	DISULFID	499	525	
DR	InterPro; IPR001117; Cu-oxidase.		FT	DISULFID	607	688	
DR	InterPro; IPR008972; Cupredoxin.		FT	DISULFID	1712	1738	
DR	InterPro; IPR000421; FA58 C.		FT	DISULFID	1894	2048	
DR	InterPro; IPR008979; Gal_bind_like.		FT	DISULFID	2053	2208	
DR	Pfam; PF00394; Cu-oxidase; 3.		FT	MOD_RES	697	697	PROBABLE.
DR	Pfam; PF00754; F5/F8 type_C; 2.		FT	MOD_RES	701	701	BY SIMILARITY.
DR	SMART; SM00231; FA58C; 2.		FT	MOD_RES	730	730	BY SIMILARITY.
DR	PROSITE; PS01285; FA58C 1; 2.		FT	MOD_RES	1513	1513	SULFATION (POTENTIAL).
DR	PROSITE; PS01286; FA58C 2; 2.		FT	MOD_RES	1529	1529	SULFATION (POTENTIAL).
DR	PROSITE; PS50022; FA58C 3; 2.		FT	MOD_RES	1537	1537	SULFATION (POTENTIAL).
DR	PROSITE; PS00079; MULTICOPPER_OXIDASE1; 2.		FT	MOD_RES	1541	1541	SULFATION (POTENTIAL).
KW	Blood coagulation; Glycoprotein; Sulfation; Calcium; Signal; Zymogen;		FT	CARBOHYD	225	225	N-LINKED (GLCNAC. .)
KW	Repeat.		FT	CARBOHYD	239	239	N-LINKED (GLCNAC. .)
FT	SIGNAL.	1 28	FT	CARBOHYD	297	297	N-LINKED (GLCNAC. .)
FT	CHAIN	29 2211	FT	CARBOHYD	382	382	N-LINKED (GLCNAC. .)
FT	CHAIN	29 741	FT	CARBOHYD	460	460	N-LINKED (GLCNAC. .)
FT	PROPEP	742 1564	FT	CARBOHYD	553	553	N-LINKED (GLCNAC. .)
FT	CHAIN	1565 2211	FT	CARBOHYD	587	587	N-LINKED (GLCNAC. .)
FT	CHAIN		FT	CARBOHYD	745	745	N-LINKED (GLCNAC. .)
FT	CHAIN		FT	CARBOHYD	756	756	N-LINKED (GLCNAC. .)


```
Db 1507 SRDDGDYIEIPROKESSESDYGBFEFVAYNDPVQTDLR-----DINSRNPEDNIAA 1560
Qy 1708 ---RSFOKTHFYFAAVERLWDYGMSSSPVLRNRAQSGS---VPQ---PKKVVFPQFT 1758
Dd 1561 WYLRSTNGRKYIYAABEISWDYS-----KFGQSDVDYVPEDTVKKVFRKYL 1611
Qy 1759 DGSFTQPLRGELNHEHLGLGPIYRAEVEDNIMVTFRNOASRPYSFYSLSIYE----- 1812
Dd 1612 DSTFKLPQGEYEHLGILGEVIRAEVDDVTVQRFKNLARPISLHAGHSUYEKSSGK 1671
Qy 1813 --EDORQAEPRKNVKNETKTYPKVQHHEMPTKDEFCKAWAYFSDVDLEKDVHSL 1870
Dd 1672 TYEDDSPEWFKEDNAIQNKTYVYVHATSTRSPENPCSGACAWAYSAVNPEDKIHSL 1731
Qy 1871 IGPLLVCHTNILNPAHQVQVTVQZPALFTYDFETKSNYFTENMERNCRAPCNIOMEDPT 1930
Dd 1732 IGPLLICKKGLDKETNPVDMRFEVLFMVFDKKSWIYDKKPTRSWRRASS-----E 1785
Qy 1931 FKENVRFHAINGYIMDTLPGLVMAQDQIRWYLLSMGSENIHSTHPSGHVPTVRKBEY 1990
Dd 1786 VNSHEFHAINGMVX-IPGLRMVYQEWVRHLNLNGLGSRDIEHVHFGQILLENCTQH 1844
Qy 1991 KVALNLYPVGFETVEMLPKAGIWEVCLIGEHLHAGNSTLFLVYSNKQOTPLGMASSH 2050
Dd 1845 QLGVMPLLPKSGFKILEMKASKPGWMLLDEVEGIEIQAGWQTPFLIVDRECKMPGLSTGL 1904
Qy 2051 TRDFOITASGOYGQWAPKIALRHYSNSAW-----STK-EPFSKIKVDLLAPMI-HGIK 2104
Dd 1905 IADSQIASEFNGYWEPLKRLNNGSGSNVIAEKLSTFEPPEPWIQVDMQKEVLLTGIG 1964
Qy 2105 TOGARQKSSSLYISOFIIMYS:DGKKWQTYRGNSTGLMVFGNVDSGIGKHINFPPII 2164
Dd 1965 TOGAKHLKPYTYTEFCVAYSLDKQWRIFKGNSTRNWMYFGNSDASTIKENQIDPPV 2024
Qy 2165 ARYIRLPHTHYSIRTLRMELMGDLNCSMPLGMESKAISDAQTASSYFTNMPAT-WS 2223
Dd 2025 ARYIRISPTGTSYKPKALRELGQCEVNGCSTPLGMESGKIENKQITASSFKSKWNGYWE 2084
Qy 2224 PSKARLHLQGRNAPVQNNPKELQVDFQKTMKVTVGTQGVKSLTSMYVKBFLSS 2283
Dd 2085 PFLARLNAGRVNAWAKANNNNQWLDLLKIKKITAIVTQCKSLSEMTVKSYTHY 2144
Qy 2284 SQDGHQWILFFONGKV--KVFGNQDSFTPVVNSLDPPLTLRYLRIHPQSWVHQIALRME 2341
Dd 2145 SDQGTDWPKYREKSSWMDKIFPGNNKVRGHVKNFNPPIISFIRIIPKTNQSIARLE 2204
Qy 2342 VLGCEAQDLY 2351
Dd 2205 IFGC---DMY 2211

RESULT 7
CERU HUMAN
ID CERU HUMAN STANDARD; PRT: 1065 AA.
AC P00450; Q14063;
DT 21-JUL-1986 (Rel. 01, Created)
DT 13-AUG-1987 (Rel. 05, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Ceruloplasmin precursor (EC 1.16.3.1) (Ferroxidase).
GN CP.
OS Hmo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=86259737; PubMed=2873574;
RA Koschinsky M.L., Funk W.D., van Oost B.A., McGillivray R.T.A.;
RT "Complete cDNA sequence of human ceruloplasmin.";
RL Proc. Natl. Acad. Sci. U.S.A. 83:5086-5090(1986).
RN [2]
RP SEQUENCE OF 1-1006 FROM N.A.
```

```
RX MEDLINE=95217183; PubMed=7702601;
RA Daimon M., Yamatani K., Igarashi M., Fukase N., Kawanami T.,
RA Kato T., Tomimaga M., Sasaki H.;
RT "Fine structure of human ceruloplasmin gene.";
RL Biochem. Biophys. Res. Commun. 208:1028-1035(1995).
RN [3]
RP SEQUENCE OF 1-40; 549-599; 784-829 AND 919-952 FROM N.A.
RX MEDLINE=86275241; PubMed=3755405;
RA Mercer J.F.B., Grimes A.;
RT "Isolation of a human ceruloplasmin cDNA clone that includes the N-
RT terminal leader sequence.";
RL FEBS Lett. 203:185-190(1986).
RN [4]
RP SEQUENCE OF 218-1065 FROM N.A.
RX MEDLINE=86205876; PubMed=3486416;
RA Yang F., Naylor S.L., Lum J.B., Cutshaw S., McCombs J.J.,
RA Naberhaus K.H., McGill J.R., Adrian G.S., Moore C.M., Barnett D.R.,
RA Bowman B.H.;
RT "Characterization, mapping, and expression of the human ceruloplasmin
RT gene.";
RL Proc. Natl. Acad. Sci. U.S.A. 83:3257-3261(1986).
RN [5]
RP SEQUENCE OF 20-1065.
RX MEDLINE=84119493; PubMed=6582496;
RA Takahashi N., Ortel T.L., Putnam F.W.;
RT "Single-chain structure of human ceruloplasmin: the complete amino
RT acid sequence of the whole molecule.";
RL Proc. Natl. Acad. Sci. U.S.A. 81:390-394(1984).
RN [6]
RP SEQUENCE OF 158-333; 518-724 AND 858-1065.
RX MEDLINE=83117800; PubMed=6571985;
RA Takahashi N., Bauman R.A., Ortel T.L., Dwulet F.E., Wang C.-C.,
RA Putnam F.W.;
RT "Internal triplication in the structure of human ceruloplasmin.";
RL Proc. Natl. Acad. Sci. U.S.A. 80:115-119(1983).
RN [7]
RP SEQUENCE OF 501-905.
RX MEDLINE=81199407; PubMed=6940148;
RA Dwulet F.E., Putnam F.W.;
RT "Complete amino acid sequence of a 50,000-dalton fragment of human
RT ceruloplasmin.";
RL Proc. Natl. Acad. Sci. U.S.A. 78:790-794(1981).
RN [8]
RP SEQUENCE OF 907-1065.
RX MEDLINE=80137543; PubMed=6987229;
RA Kingston I.B., Kingston B.L., Putnam F.W.;
RT "Primary structure of a histidine-rich proteolytic fragment of human
RT ceruloplasmin. II. Amino acid sequence of the tryptic peptides.";
RL J. Biol. Chem. 255:2886-2896(1980).
RN [10]
RP SEQUENCE OF 1007-1061 FROM N.A.
RX MEDLINE=90285218; PubMed=2355023;
RA Yang F.Y., Friedrichs W.E., Cupples R.L., Banifacio M.J.,
RA Sanford J.A., Horton W.A., Cupples B.H.;
RT "Human ceruloplasmin. Tissue-specific expression of transcripts
RT produced by alternative splicing.";
RL J. Biol. Chem. 265:10780-10785(1990).
RN [11]
RP REVIEW.
RX MEDLINE=22049919; PubMed=12055353;
RA Hellman N.E., Gitlin J.D.;
RT "Ceruloplasmin metabolism and function.";
RL Annu. Rev. Nutr. 22:439-458(2002).
RN [12]
RP X-RAY CRYSTALLOGRAPHY (3.1 ANGSTROMS).
```

RA Zaitseva I., Zaitsev V., Card G., Moshkov K., Bax B., Ralph A.,
 RA Lindley P.;
 RA "The X-ray structure of human serum ceruloplasmin at 3.1 Å: nature of
 RT the copper centres.";
 RL J. Biol. Inorg. Chem. 1:15-23(1996).
 CC !- FUNCTION: Ceruloplasmin is a blue, copper-binding (6-7 atoms per
 CC molecule) glycoprotein found in plasma. Four possible functions
 CC are ferroxidase activity, amine oxidase activity, copper transport
 CC and homeostasis, and superoxide dismutase activity.
 CC !- CATALYTIC ACTIVITY: 4 Fe(2+) + 4 H(+) + O(2) = 4 Fe(3+) + 2 H(2)O.
 CC !- COFACTOR: Binds 6 Cu-ions per molecule. This protein belongs to
 CC the multicopper oxidases which contain three distinct Cu centers
 CC known as type 1 or blue, type 2 or normal, and type 3 or coupled
 CC binuclear
 CC !- TISSUE SPECIFICITY: Expressed by the liver and secreted in plasma.
 CC !- DISEASE: Defects in CP are the cause of aceruloplasminemia
 CC [MIM:604290]. It is an autosomal recessive disorder of iron
 CC metabolism characterized by iron accumulation in the brain as well
 CC as visceral organs. Clinical features consist of the triad of
 CC retinal degeneration, diabetes mellitus and neurological
 CC disturbances.
 CC !- DISEASE: Ceruloplasmin levels are decreased in Wilson's disease,
 CC in which copper cannot be incorporated into ceruloplasmin in
 CC liver because of defects in the copper-transporting ATPase 2.
 CC !- SIMILARITY: Contains 3 F5/8 type A domains.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 DR EMBL; M13699; AAA51976.1; -;
 DR EMBL; D45045; BAA08085.1; -;
 DR EMBL; D45044; BAA08084.1; -;
 DR EMBL; D45028; BAA08084.1; JOINED.
 DR EMBL; D45029; BAA08084.1; JOINED.
 DR EMBL; D45030; BAA08084.1; JOINED.
 DR EMBL; D45031; BAA08084.1; JOINED.
 DR EMBL; D45032; BAA08084.1; JOINED.
 DR EMBL; D45033; BAA08084.1; JOINED.
 DR EMBL; D45034; BAA08084.1; JOINED.
 DR EMBL; D45035; BAA08084.1; JOINED.
 DR EMBL; D45036; BAA08084.1; JOINED.
 DR EMBL; D45037; BAA08084.1; JOINED.
 DR EMBL; D45038; BAA08084.1; JOINED.
 DR EMBL; D45039; BAA08084.1; JOINED.
 DR EMBL; D45040; BAA08084.1; JOINED.
 DR EMBL; D45041; BAA08084.1; JOINED.
 DR EMBL; D45042; BAA08084.1; JOINED.
 DR EMBL; D45043; BAA08084.1; JOINED.
 DR EMBL; D45044; BAA08084.1; JOINED.
 DR EMBL; D45045; BAA08084.1; JOINED.
 DR EMBL; X04135; CAA27752.1; -;
 DR EMBL; X04136; CAA27753.1; -;
 DR EMBL; X04137; CAA27754.1; -;
 DR EMBL; X04138; CAA27755.1; -;
 DR EMBL; M13536; CAA51975.1; -;
 DR EMBL; J05506; -; NOT_ANNOTATED_CDS.
 DR PIR; A25443; KUHU.
 DR PDB; 1KCW; 12-FEB-97.
 DR GlycoSuiteDB; P00450; -;
 DR SWISS-2DPAGE; P00450; HUMAN.
 DR Siera-2DPAGE; F00450; -;
 DR GeneW; HGNC:12295; CP.
 DR MIM; 117700; -;
 DR MIM; 604290; -;
 DR GO; GO:0005615; C:extracellular space; TAS.
 DR GO; GO:0004322; F:ferroxidase activity; TAS.
 DR GO; GO:0006879; P:iron ion homeostasis; TAS.
 DR InterPro; IPR001117; Cu-oxidase.
 DR InterPro; IPR002355; Cu_ox_copper_Es.

DR InterPro; IPR008972; Cupredoxin.
 DR Pfam; PF00394; Cu-oxidase; 3.
 DR PROSITE; PS00079; MULTICOPPER_OXIDASE1; 3.
 DR PROSITE; PS00080; MULTICOPPER_OXIDASE2; 1.
 KW Transport; Ion transport; Copper transport; Oxidoreductase; Plasma;
 KW Metal-binding; Copper; Repeat; Signal; Glycoprotein; Polymorphism;
 KW 3D-structure.
 FT SIGNAL 1 19 CERULOPLASMIN.
 FT CHAIN 20 1065 F5/8 TYPE A 1.
 FT DOMAIN 20 357 PLASTOCYANIN-LIKE 1.
 FT DOMAIN 20 200 PLASTOCYANIN-LIKE 2.
 FT DOMAIN 209 357 PLASTOCYANIN-LIKE 2.
 FT DOMAIN 370 718 F5/8 TYPE A 2.
 FT DOMAIN 370 560 PLASTOCYANIN-LIKE 3.
 FT DOMAIN 570 718 PLASTOCYANIN-LIKE 4.
 FT DOMAIN 730 1061 F5/8 TYPE A 3.
 FT DOMAIN 730 900 PLASTOCYANIN-LIKE 5.
 FT DOMAIN 908 1061 PLASTOCYANIN-LIKE 6.
 FT CARBOHYD 138 138 N-LINKED (GLCNAC. .).
 FT CARBOHYD 358 358 N-LINKED (GLCNAC. .).
 FT CARBOHYD 357 397 N-LINKED (GLCNAC. .).
 FT CARBOHYD 762 762 N-LINKED (GLCNAC. .).
 FT DISULFID 174 200 PROBABLE.
 FT DISULFID 276 357 PROBABLE.
 FT DISULFID 534 560 PROBABLE.
 FT DISULFID 637 718 PROBABLE.
 FT DISULFID 874 900 PROBABLE.
 FT METAL 120 120 COPPER (TYPE 2) (BY SIMILARITY).
 FT METAL 122 122 COPPER (TYPE 3) (BY SIMILARITY).
 FT METAL 180 180 COPPER (TYPE 3) (BY SIMILARITY).
 FT METAL 182 182 COPPER (TYPE 3) (BY SIMILARITY).
 FT METAL 994 994 COPPER (TYPE 1) (BY SIMILARITY).
 FT METAL 997 997 COPPER (TYPE 2) (BY SIMILARITY).
 FT METAL 999 999 COPPER (TYPE 3) (BY SIMILARITY).
 FT METAL 1039 1039 COPPER (TYPE 3) (BY SIMILARITY).
 Query Match 10.6%; Score 1322; DB 1; Length 1065;
 Best Local Similarity 19.2%; Pred. No. 1.3e-60;
 Matches 398; Conservative 203; Mismatches 417; Indels 1054; Gaps 29;
 QY 5 LSTQFFLCRLRCFSAIRRYLGVAVELSWDMQSDLGE---LPVDARPEPRVPKFPFNT 61
 DB 6 LGIFLFLCSTP-AWAKEKHYVIGITETTWDY-ASDHGEKKJISVDESHNIYQNGPDI 63
 QY 62 SVVYKTLFVEPTDHLNIAKPRPMGLIGTIOAEVYDTVTILKMAHPVSLHVG 121
 DB 64 GSDYKRAJYLQYDTEFTTTEKPVMLGFLGLIIKAETGDKVYVHLKMLASPYTFHSHG 123
 QY 122 VSYWKASGAEYDDOTSOREKEDDKVFFGGSHTYVWQVLKENGPRVASDPLCLTYSLSHV 181
 DB 124 ITYKHEGAIYPDNTDFQRAADKVPFGEOVTYMLLATEEOSPGEGDNCVTRIYHSHI 183
 QY 182 DLVKNLSGLIGALLVCREGSLAKTKTQL-HKFIILFAVDFEGKSWHSE-----TKNS 234
 DB 184 DAPKDIAAGLIQPLIICKKDSLDKEKHIIDREFVYMFVSVVDENFSWLEKNIKTYCSEP 243
 QY 235 LMQDRDAASARAWPKWHTVNGVWNSLPGLICCHRSKSVVHVHVGITPTPEVHSILEGHT 294
 DB 244 EKVDKDNEDFQESNRMYISVNGTTFGLSPGLSCADRDVKVYLFMGNEVDVHAATFHQA 303
 QY 295 FLVRNHRQASLEISPTITLTAOTLLMDLQFLLSCHISSHQHDGMEAYVYKVDSCPEEPQL 354
 DB 304 LTNKYVRIDTINLFFATLFDAYWVAQNPGEWMLSCQNLNHLKAGLQAFQVQEC----- 357
 QY 355 RMKNNEBAEDYDDDLTDSEMDVVRDDDNSPSFIOIRSVAKKHPTWVHYIAAEEDWDY 414
 DB 358 ---NKSSSKD-----NIRCKHVRH-----VYIAAEETIWNV 365
 QY 415 APL-----VLAPDORSYKSYQLNNGPDRIGKYYKVRFMAYTDTF---KTREATQ 462
 DB 386 ABSGIDIFTKENLITAPGSDS---AVFEQGTTRIGGSYKKLVYREYTDASTNKRGRGEE 443
 QY 463 HESGILGPLLYGEVGDTLIIIFKNOASRPNIYPHGI-----TDVRLPY---SRRLP 511

```

Db 444 EHLGILGPVIAEVDGDIRTFHFKGAYPLSIEPIGVRFNKNNEGTVYSNYPQSRSP 503
QY 512 KGVEHLKDFILPGEIFKXKWTYVDGPTKSPRCGLTRYSSFYNNKRDASLGLPULL 571
Db 504 PSASH-----VAFETFTVETWTVKXVGPNDPVCCLAKMYSAVDFTKDIFTGLIGPMK 558
QY 572 ICYKESVDORGQIMSDKRNVLFSVFDENRSWLTJENIQRFLPNAGVQLEDPEFOASN 631
Db 559 ICKKGSILHANGROKQVDKRFYLPFTTFDENESILLLEDNIEMFTTADQVDKEDDEQBSN 618
QY 632 IMESINGYVDSIQ-LSVCLHEVAYWYLLSIGAQDTFLSVFFSGYTFKXKMYVEDTLTLF 690
Db 619 KMFESXNGFMYGNQPGITMCKGDSVWVYLFSAAGNEADVHGIFYFSGNTYLRGRERDANLF 678
QY 691 PFSGETVFSMENPGLWILLGCHNSDFRNGMTALLKVVSSCKNTGDIYEDSYDIDISAYLL 750
Db 679 PQTSILEWPDTEGTFNVECLTIDHYTGKMKQKTYVNOQ----- 718
QY 751 SKNNAIEPRFSQNSRHPSTRQKFNATTIPENDIEKTPWFAHRTMPKIQNVSSDDL 810
Db 719 ----- 718
QY 811 MILLRQSPTHGLSLQAKYETPSDDPSGAIDNNSLSNTHFRPOLHSHGDMVFTF 870
Db 719 --RROSE-----DSTFYLGERTYY----- 735
QY 871 ESGQLRLNEKLTGTTAATLKKLDFKVSSTSNLLISTIPSDNIAAGTDNTSSLGPPSPMV 930
Db 736 -----IAAVEVE----- 742
QY 931 HYDSQDHTTFGKSSPLTESGGPLSLSEENDSKILLESGLMNSQBSWGKVSSTESGR 990
Db 743 ----- 742
QY 991 LFKGKEAGPALLTKDNALPKVSIISLTKNTKNSNATNRKTHIEDGPSLLIENSPSVQON 1050
Db 743 -----KD- 744
QY 1051 ILESDETFKKTPLIHDRYMLDKNATALNLHNSKNTTSKKNVWQKKEGPIPPDAQN 1110
Db 745 ----- 744
QY 1111 PDMSPFKMLFLPESARWIQRTGKNSLNSGQSPSKQLVSLGPEKSVQGNFLSEKKNV 1170
Db 745 ----- 744
QY 1171 VGKGFTKDVGLKXWVFPSSRNILFLTNLDNLHENNTHTNOEKKIQEEIEKKTLEIQENVVL 1230
Db 745 ----- 744
QY 1231 POIHTVTGTFKNFMKNLFLILSTRQNEGSYDGAYAPVLQDFRSLNDSTNRTKHTAHFSKK 1290
Db 745 -----YSP----- 747
QY 1291 GREENLEGIGNOTKOIVEKYACTTRISPNTSQNFVQRSKALKQFRLPLEBETLEKRI 1350
Db 748 -----QR----- 749
QY 1351 IVDDTSTOWSKNMKELTPSTLTQIDYNEKEKAITQSPLSDCLTTRSHSIPQANRSPLPDIA 1410
Db 750 -----EWEKELHL----- 758
QY 1411 KVSSPFSIRPIYLRVLFDONSSHLPAASRYKKDSGVQESSHFLQAKKNLSLAILTLE 1470
Db 759 -----GEQVSNAPL----- 768
QY 1471 MTGDQREVGSLGTSATNSVTYKXVENTVLPKDLPKTSKVELLPKHVIYQKDLFPPTETS 1530
Db 769 ----- 768
QY 1531 NGSPGHLDIVBGSLLQGTGALKNEANRPGKVPFLRVATPSSAKTPSKLLDPLAWDNHY 1590

```

```

Db 769 ----- 768
QY 1591 GTQIPKEWKSORKSEKTAFAKKOTILSNACESHAIAINEGQMKPEIEVTWAKOGR 1650
Db 769 ----- 768
QY 1651 TERLCSQNPVLK3HOREITRITLQSDQBEIDYDDTISVEMKKEDEFDIYDEDENQSPRF 1710
Db 769 -----DKGE----- 772
QY 1711 OKKTRHYFIAAVERLMDYGMSSSPHVLNRNRAQSGSVPOFKKVVFOEFTDGSFTQPLYRGE 1770
Db 773 -----FYIGS-----KXKVVYRCYTDTSTFRVPVERKA 800
QY 1771 LNEHLGLICPYRAEVEDNIMVTFRNOASRPYSFYSSLSISYEDORQGAEPKRNFKV--- 1827
Db 801 EEEHLGILGPQHADVGDKVJIFKMATRPSYIHA-----HGVQTSSTVTPTL 850
QY 1828 PNETKTYFMKVQHMAPTKDBFDCAKAWYPSVDLEKDVHSGLIGPLLKYCHTINTLPAHG 1887
Db 851 PGETITVYWKIPERGAGTEDSACIPWAYYSTVDQVKDLYSLGLIGPLIVCRPPLYKVFNP 910
QY 1888 ROVTYQOEAFALFTTIFDETKSWYFTENMERNCRAPCNIOMEDPTFKENYRFAHNGYIMDT 1947
Db 911 RRKL--EPALFLVFDENESWLDONIKTYSDBPEKWKDDREFTESNKMWHAINGMFCN 968
QY 1948 LFLGVYAQDQRTWLLSMGNSNENIHSFSGHVFTVRKKEBYKMYLNLYPGVFETVM 2007
Db 969 LQGLTMHVGDYENVYLMGMNEIDLHTVHFHGSFYKHRGVYSSVDVDFIDPGTYQTLEM 1028
QY 2008 LPSKAGIWRVECLIGEHLHAGMSTLFLVYSNK 2039
Db 1029 PRFTGIMLLHCHVTDHIIHAGMETTYTVLQNE 1060

RESULT 8
CERU RAT
ID CERU RAT STANDARD; PRG; 1059 AA.
AC P13535; Q64719;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Ceruloplasmin precursor (EC 1.16.3.1) (ferroxidase).
GN Cp.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Sprague-Dawley; TISSUE=Liver, and Lung;
RX MEDLINE=90237081; PubMed=2332446;
RA Fleming R.E., Gitlin J.D.;
RT "Primary structure of rat ceruloplasmin and analysis of
RT tissue-specific gene expression during development.";
RL J. Biol. Chem. 265:7701-7707(1990).
RN [2]
RP SEQUENCE OF 257-294; 571-612 AND 823-892 FROM N.A.
RC TISSUE=Liver;
RX MEDLINE=87137545; PubMed=3818625;
RA Aldred A.R., Grimes A., Schreiber G., Mercer J.F.B.;
RT "Rat ceruloplasmin. Molecular cloning and gene expression in liver,
RT choroid plexus, yolk sac, placenta, and testis.";
RL J. Biol. Chem. 262:2875-2878(1987).
CC -!- FUNCTION: Ceruloplasmin is a blue, copper-binding (6-7 atoms per
CC molecule) glycoprotein found in plasma. Four possible functions
CC are ferroxidase activity, amine oxidase activity, copper transport
CC and homeostasis, and superoxide dismutase activity.
CC -!- FUNCTION: May also play a role in fetal lung development or
CC pulmonary antioxidant defense.
CC -!- CATALYTIC ACTIVITY: 4 Fe(2+) + 4 H(+) + O(2) = 4 Fe(3+) + 2 H(2)O.
CC -!- COFACTOR: Binds 6 Cu-ions per molecule. This protein belongs to
CC the multicopper oxidases which contain three distinct Cu centers

```



```
Qy 1057 EFKQVTELDHRLMLNDKRNATRLNHNKNTSSKNMEMVOQKEGPPDPAQNPDMSEFF 1116
Db 733 ----- 732
Qy 1117 KMLFLPESARWQTHGKNSLNSGGPSPKQVLSIGPEKSVGQNFLEKKNVVGKEF 1176
Db 733 ----- 732
Qy 1177 TKVGLKEMVPPSERNLFLTKDLNHNHNTHNOEKKIOEEIKKETTLOENVLPQIHV 1236
Db 733 ----- 732
Qy 1237 TGTNFMKNLFLSTRQNVESYDCAVAVLQDFRSLNDSNTRKKTHTAHSKKGEENL 1296
Db 733 ----- 741
Qy 1297 ECLGNTQKQIVEKVACTTRISPTNSQNFVQRSKRLKQFRLPLETELEKRIIVDDTS 1356
Db 742 ----- 742
Qy 1357 TQSKQMKHLPTSLTQIDYNEKEKGAITQSPCLSTRSHSIPQANRSPPIAKVSGFP 1416
Db 743 RDEWEMELFHL ----- 752
Qy 1417 SIRPIYLRVLFQDNSSHLPAASYEKDQVQESSHPLOQAKKNLSLAILTEMTGQR 1476
Db 753 ----- 762
Qy 1477 EVGSLGTSATNSVTKYKVENTVLPKPDLPKTSKGVKELLPKVHYQKDLFPPTETSGSPQH 1536
Db 763 ----- 762
Qy 1537 LDLVEGSLLOQTEGAIKNEANRPGKPLRVATESSAKTSPKLLDPLANDHNYGTQPK 1596
Db 763 ----- 762
Qy 1597 ZEWKQKSPKTAKKKDTILSNACSNHAIAINNEGQNKPEIEVTWAKGRTERLCS 1656
Db 763 ----- 762
Qy 1657 QNPPVLKRQREITRTLQSDQEHYDDDTISVEMKKEFDIYDEQSPRSQKQTRH 1716
Db 766 ----- 766
Qy 1717 VPIAAVERLWDYGMSSPHVLNRAQSGVPQKVVFOEPTDGSFTOPLYRGELNEHLG 1776
Db 767 FPIGS ----- KYKKVYVREFTDSTFREQVKERABEEHLG 800
Qy 1777 LAGPVIRAEVEDNIMVTFERNASRPYSFYSLLSYEEDQOQAEPRKQVK--FNETKT 1833
Db 801 MLGPIHLHADVGAQKVVFKNMATRPYSIHA-----HGVKTKSSVAPTLGCEVKT 850
Qy 1834 YFWQVQHMAPTKDEFDCKAWAYFSDVLEKDVHSLGILGLLVC---HTNTLNPAHGRQV 1890
Db 851 YLWQIPERSGAGTDSPCIPWAYSVNDRVKDLYSLGLIVCRKSYKVNPK----- 905
Qy 1891 TVQEFALFTIPEETKSVYFTNMERNCRAPCNIOHEDPTFKNVFRFAINGYMDTLP 1950
Db 906 KXVEFSLLFLVDFENESWYLDNINITYDPHPKDKNDNEEFESKKNHAIINGKMFNLQ 965
Qy 1951 LVMAQQRIRWYLLSGNSNENIHSIFSGHVFTRVKKEBYKVALNLYPGVFETVEMLP 2010
Db 966 LTHVGVDEWVWVMANGNEIDLHTVHFHGHSSQYKHRGIHSSDVFDFPCTQTLEMEFQ 1025
Qy 2011 KAGIWEVECLIGELHAGMSTLFLVYSNK 2039
Db 1026 TSGTWLHCHVTDHAGWVTVTVLPNQ 1054
PRT; 1062 AA.
```

RESULT 9
CERU_MOUSE

ID_CERU_MOUSE

```
AC 061147;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Ceruloplasmin precursor (EC 1.16.3.1) (Ferroxidase).
GN CP.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RN [2]
RA Klomp L.W.J., Farhangrazi Z.S., Choi D.W., Gitlin J.D.;
RL Submitted (MAR-1996) to the EMBL/GenBank/DBJ databases.
RP TISSUE SPECIFICITY.
RX MEDLINE=96294736; PubMed=8690795;
RA Klomp L.W.J., Farhangrazi Z.S., Dugan L.L., Gitlin J.D.;
RL "Ceruloplasmin gene expression in the murine central nervous system.";
J. Clin. Invest. 98:207-215(1996).
CC -!- FUNCTION: Ceruloplasmin is a blue, copper-binding (6-7 atoms per
CC molecule) glycoprotein found in plasma. Four possible functions
CC are ferroxidase activity, amine oxidase activity, copper transport
CC and homeostasis, and superoxide dismutase activity.
CC -!- CATALYTIC ACTIVITY: 4 Fe(2+) + 4 H(+) + O(2) = 4 Fe(3+) + 2 H(2)O.
CC -!- COFACTOR: Binds 6 Cu-ions per molecule. This protein belongs to
CC the multicopper oxidases which contain three distinct Cu centers
CC known as type 1 or blue, type 2 or normal, and type 3 or coupled
CC binuclear.
CC -!- TISSUE SPECIFICITY: Expressed in many tissues, including liver,
CC eye and brain.
CC -!- SIMILARITY: Contains 3 F5/8 type A domains.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U49430; AAB07596.1; -.
CC HSP; P0450; 1KCW.
CC GSD; MGI:88476; Co.
CC InterPro; IPR001117; Cu-oxidase.
CC InterPro; IPR002355; Cu_ox_copper_BS.
CC InterPro; IPR008972; Cupredoxin.
CC Pfam; PF00394; Cu-oxidase; 3.
CC PROSITE; PS00079; MULTICOPPER_OXIDASE1; 3.
CC PROSITE; PS00080; MULTICOPPER_OXIDASE2; 1.
CC Transport; Ion transport; Copper transport; Oxidoreductase; Plasma;
CC Metal-binding; Copper; Repeat; Signal; Glycoprotein.
CC SIGNAL 1 19
CC CHAIN 20 1062
CC CERULOPLASMIN.
CC F5/8 TYPE A 1.
CC DOMAIN 20 199
CC PLASTOCYANIN-LIKE 1.
CC DOMAIN 208 356
CC PLASTOCYANIN-LIKE 2.
CC DOMAIN 369 713
CC F5/8 TYPE A 2.
CC DOMAIN 369 555
CC PLASTOCYANIN-LIKE 3.
CC DOMAIN 565 713
CC PLASTOCYANIN-LIKE 4.
CC DOMAIN 725 1057
CC F5/8 TYPE A 3.
CC DOMAIN 725 896
CC PLASTOCYANIN-LIKE 5.
CC DOMAIN 904 1057
CC PLASTOCYANIN-LIKE 6.
CC F5/8 TYPE A 4.
CC DISULFID 173 199
CC BY SIMILARITY.
CC DISULFID 275 356
CC BY SIMILARITY.
CC DISULFID 529 555
CC BY SIMILARITY.
CC DISULFID 632 713
CC BY SIMILARITY.
CC DISULFID 870 896
CC BY SIMILARITY.
CC METAL 120 120
CC COPPER (TYPE 2) (BY SIMILARITY).
CC METAL 122 122
CC COPPER (TYPE 3) (BY SIMILARITY).
CC METAL 179 179
CC COPPER (TYPE 3) (BY SIMILARITY).
CC METAL 181 181
CC COPPER (TYPE 3) (BY SIMILARITY).
CC METAL 990 990
CC COPPER (TYPE 1) (BY SIMILARITY).
```

Query Match 10.0%; Score 1242; DB 1; Length 1062;
 Best Local Similarity 18.9%; Pred. No. 1.7e-56;
 Matches 393; Conservative 196; Mismatches 432; Indels 1054; Gaps 30;

```

FT METAL 993 COPPER (TYPE 2) (BY SIMILARITY) .
FT YETAL 995 COPPER (TYPE 3) (BY SIMILARITY) .
FT METAL 1036 COPPER (TYPE 3) (BY SIMILARITY) .
FT METAL 1036 COPPER (TYPE 1) (BY SIMILARITY) .
FT METAL 1037 COPPER (TYPE 3) (BY SIMILARITY) .
FT METAL 1041 COPPER (TYPE 1) (BY SIMILARITY) .
FT METAL 1046 COPPER (TYPE 1) (BY SIMILARITY) .
FT CARBOHYD 138 N-LINKED (GLCNAC. .) (POTENTIAL) .
FT CARBOHYD 226 N-LINKED (GLCNAC. .) (POTENTIAL) .
FT CARBOHYD 583 N-LINKED (GLCNAC. .) (POTENTIAL) .
FT CARBOHYD 625 N-LINKED (GLCNAC. .) (POTENTIAL) .
FT CARBOHYD 757 N-LINKED (GLCNAC. .) (POTENTIAL) .
FT CARBOHYD 922 N-LINKED (GLCNAC. .) (POTENTIAL) .
SQ SEQUENCE 1062 AA; F3F52ED09A238F16 CRC64;

```

Query Match 10.0%; Score 1242; DB 1; Length 1062;
 Best Local Similarity 18.9%; Pred. No. 1.7e-56;
 Matches 393; Conservative 196; Mismatches 432; Indels 1054; Gaps 30;

```

QY 10 FLCLLRFCE-----SATERVYLGAVELSWDYMQ--SDLGELPVDARPPRPVPSFFPN 60
DB 3 FLLLSFFLYSSALARDKHVFIGITEAVWYASCTEEKKLISVDTEQSNFYLQNGPDR 52
QY 61 TSVVYKTLFVFTDHLFNIARPPWMLGELPTIOAENVYDVVITLKNWASHPVSLHAV 120
DB 63 IGRKYKALYFEYDGTGTSKTDKPAWGLFGLPVKAIVEDKYVYHLKNLASRIYTFHAH 122
QY 121 GVSVMKASGEAVDDOTSOREDDKVPFGGSHYVMVLKENGPMASDPLCLTYSLSH 180
DB 123 GVITYKEYGAVPNDTDFQADDKVLFGQQVYVYLHA-NPSPGEGDSNCVTRIYHSH 181
QY 181 VDVKDLNGLIGALLVCREGLAKEKTQTL-HKF-LLPVAFDEGKSWSETKNSLM--- 236
DB 182 VDAPXDIASGLIGLILCKGSLYKEKERNIDQEFVIMPVVDENLSWLEDNIRKTFQSE 241
QY 237 ---QBRDAASAWPKGTWNGVNRSLPGLICCHRKSYVWVIGMTTPEVHSIFLEGH 293
DB 242 PKVKDXNEDFQESNRMYINGVTFGSLPGLNCAADRVKWLFGMGNEVDVESAFPHGQ 301
QY 294 TPLVRNHRQASLEISPIITLTAQTLMDLGOFLLSCHISSHQHGWMEAYVKVYDSCPEEPQ 353
DB 302 ALTSRNYQTDIINLFFATLIDAYMAQNGVWMLSCQNLHLKAGLQAFQVQDC----- 356
QY 354 LRKNNNEEADYDDJTDSEMDVVRDDNSPFIQIRSVAKKHPTWVHYIAAEEEDWD 413
DB 357 ---NKPSKDK-----NIRKGVHREH-----YIAABEVIWN 383
QY 414 YAP--LVLPAPDR-----SYKSOYLNNGPQRIGRKYKURFMAYTDETP---KTREAIQH 463
DB 384 YAPSGIDITFEKLTASGSDSGVFFQGATRIGGSYKQWARYEYTDGSPFNKRKRGPDSE 443
QY 464 ESGILGPLYGEYGDPTLLIIFKNQASRPNIYPHGIT---DVRPLYSRRLPKGVKHLXD 519
DB 444 HLGILGPVIAEWGVDITIKVTENKGOHLSIQPMGVSFPAENEGTYG---PPGASSQQA 500
QY 520 FP-IILPCEITFKYKWTVTVEDGPTSDPRCLTRYSSFFVNMERDLASGLIGPLLICVKSIV 578
DB 501 AGHVAPKXTTFTYEWTVPKMGPTIADPVCLSKMYGAVDPTKIDFTGLIGPMKICKKGL 560
QY 579 DQNGNOLMSDKRNVLLFSVFDENRSYLTENIQRLPNPAGVLEDPPEFQASNMHSING 638
DB 561 LADGRQKQVDEKFFLPFTVFDNESLILLDDNIRWFTHAPDQVQDKEDEDQESKNKHSNG 620
QY 639 YVFDLSQL-SVCLHEVAYWYILSIGAQDTFLSVFFSGYTFKHXVYEDTLTLPFFSGEIV 697
DB 621 FMYGNQSWPHMCLGESIVWYLPISAGNEADVHGIFYPSGNTYLCKGEEDTANLFPKXSLTL 680
QY 698 FNSMENPGWILCHNSDENRGMATALLKVSSCDKNTGYDEYSYEDIDGAYLLSKNNATE 757
DB 681 LAMPDTKGTDFVBECLTDTHTYTGGMQKQYTVNQCOR-----QEBDFTVYL----- 724
QY 758 PRFSQNSRHESFQKQFNATTIPENDIEKTDQWFAHRTMPKPIQNVSSSDJLMLLRQSP 817

```

```

DB 725 ----- 724
QY 818 TEHGLSLSLQAEKYETESDDPSGPAIDSNNSLSEMTFRPOLHSGDMVFTPESGLQIR 877
DB 725 ----- 724
QY 878 LNEKLGTTAAETKLKDFKVSSTSNLLSTIPSDNLAAGTDNTSSLGPPSPMVHYDSOLD 937
DB 725 ----- 724
QY 938 TTLFOKKSPLTESGGPLSELSEENNDKLLBSGLMNSQSSWGNVSVSTESGRLEPKGRA 997
DB 725 -----GERT 728
QY 998 HGPALLTKDONALFKVSIULLKTKNTSNNKTHIDGPSLLIENSFSVWONILLESOTE 1057
DB 729 Y-----YVDA----- 733
QY 1058 FKKVTPLIHDMLDKATALRLNHSNKTTSKKNEMVYQKKEGPIPPDAQNPDMSFPK 1117
DB 734 ----- 733
QY 1118 MLFELPESARWORTHGKNSLNSGGQPSKQLVSLGPEKSVBEGQNFSEKNKVVVGKGF 1177
DB 734 ----- 733
QY 1178 KDVGLKEMVFPSSRNLFELTNLDNLHENNTHNQEKKIQBEIEKKETLIQBNVVLPCIHTVT 1237
DB 734 ----- 733
QY 1238 GTKNFMKNLFLSTRQNVESYDGAYAPVLQDPRSLNDSTNRKHTAHFASKKGBEENLE 1297
DB 734 -----YSP----- 742
QY 1298 GLGNQTKQIVEKYACTTRISPNISQONFVTOQSKKALKQFRLPLETELEKRIIYDDTIST 1357
DB 743 -----SR 744
QY 1358 QWSKMKHLLTPSTLTQIDYNEKEKGAIQTSPSLSDCLTRSHSIPQANRSPPLPIAKVSSPS 1417
DB 745 AWEKELHHL----- 753
QY 1418 IRPIYLTVRLFQDNSSHLPAASRYKKDSGVORSHFLQAGKNNLSLAILTLEMTGDORE 1477
DB 754 -----QONVS----- 759
QY 1478 VCSLGSATNSVYKKVENTVLPKPDLPKTSKVKVLLPKVHIYQKDLPTETSNSPGHL 1537
DB 760 ----- 759
QY 1538 DLVEGSLAQGTGATKWNANRPGKVPFLRVATSSAKTPSKLLDPLAWDNHYGTQIPKE 1597
DB 760 ----- 759
QY 1598 EWSKSEKPEKTAFAKKKDTILSLNACENHAIALINEQONKPEIEVTWAKQGRTERLCSQ 1657
DB 760 ----- 759
QY 1658 NPPVLKRHQREITRITLOSQBEIDYDDTISYEMKKEDEFDIYDEDESNQSPRSFQKTRHY 1717
DB 760 -----NVFLDKBEFFI----- 770
QY 1718 FIAAVERLWDYGMSSSPHVLNRNRAQSGVPQPKVYVQFETDGSFTQPL-YRGELNEHLG 1776
DB 771 -----GS--KYKKVYVYQFTDSSPREQVKGRAABEHLG 802
QY 1777 ILGPVIRAEVDNIMVTFRNQASRPYSFYSLSLISYBEOQROGAEPKRV---KNEKNT 1833
DB 803 ILGPPPIHANVGDVKVVFKNWATREYSIHA-----HGVKTESSTVVTPLGEVAT 852
QY 1834 YPWKQCHHMAPKDEFDCKAWAYFSDVLEKDVHSLGLIPLAVC---HTNNTNPAHQV 1890
DB 853 YTWQIPERSGAGRESACIPWAYYSTVDRKDLVSLGLIPLIVCKRSYKVPSPK----- 907

```

```

QY 1891 TVQEFALFTTIDETKSWYFTENMEBNCRAPCNQIOWEDPTEKYNRPHAINGVINDTLPG 1950
Db 908 KCEFFELLFLVDENESWILDNNIKITYSHPKVKNQDNKEEFLESNKQWHAINGKMFQNLQ 967
QY 1951 LVNAQQRIRWYLLSGNSNENIHSFSGHVFTRKKEEYKMAIYNLYPGVETVEMLPS 2010
Db 968 LHMVKQDVNWMYLMGNEIDLHTVETFGHSGFOYKRGVYSSDVPFLFGTGTQVLEMFQ 1027
QY 2011 XGKINEVECLIGELHAGNSTFLFLVYSNKCQPLG 2045
Db 1028 TFGTDLHLCHVTCHVHAGMAITTVLVEQETKSG 1062

RESULT 10
MFGM_RAT
ID MFGM_RAT STANDARD; PRT; 427 AA.
AC P70490;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Lactadherin precursor (Milk fat globule-EGF factor 8) (MFG-B8) (O-
acetyl GD3 ganglioside synthase) (AGS) (MFGM).
GN MFG-B8 OR AGS.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=96374422; PubMed=8780713;
RA Ogura K., Nara K., Watanabe Y., Kohno K., Tai T., Sanai Y.;
RT "Cloning and expression of cDNA for O-acetylation of GD3
ganglioside.";
RL Biochem. Biophys. Res. Commun. 225:932-938(1996).
CC -!- FUNCTION: May be involved in phospholipid binding. Seems to
participate in the O-acetylation of GD3 ganglioside sialic acid.
CC -!- SUBCELLULAR LOCATION: Peripheral membrane protein.
CC -!- TISSUE SPECIFICITY: Spleen, lung, heart, brain and muscle.
CC -!- SIMILARITY: Contains 2 EGF-like domains.
CC -!- SIMILARITY: Contains 2 F5/8 type C domains.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC
CC -----
DR EMBL; D84068; BAAL2210.1; -.
DR F019; JC4915; J04915.
DR HSP; P00740; IEDM.
DR InterPro; IPR006209; EGF like.
DR InterPro; IPR000421; FAS6 C.
DR InterPro; IPR008979; Gal bind like.
DR InterPro; IPR006210; IEGF.
DR Pfam; PF00008; EGF; 2.
DR Pfam; PF00754; F5_F8_type_C; 2.
DR SMART; SM00181; EGF; 2.
DR SMART; SM00231; FAS6C; 2.
DR PROSITE; PS00022; EGF 1; 2.
DR PROSITE; PS01186; EGF 2; 2.
DR PROSITE; PS00026; EGF 3; 2.
DR PROSITE; PS01285; FAS6C_1; 2.
DR PROSITE; PS01286; FAS6C_2; 2.
DR PROSITE; PS00022; FAS6C_3; 2.
DR PROSITE; PS00022; FAS6C_3; 2.
KW signal; Glycoprotein; Repeat; EGF-like domain; Milk.
FT SIGNAL 1 22 POTENTIAL.
FT CHAIN 23 427 LACTADHERIN.
FT DOMAIN 24 61 EGF-LIKE 1.
FT DOMAIN 64 108 EGF-LIKE 2.

```

```

FT DOMAIN 111 267 F5/8 TYPE C 1.
PT DOMAIN 272 427 F5/8 TYPE C 2.
FT DISULFID 28 39 BY SIMILARITY.
FT DISULFID 33 49 BY SIMILARITY.
FT DISULFID 51 60 BY SIMILARITY.
FT DISULFID 68 79 BY SIMILARITY.
FT DISULFID 73 96 BY SIMILARITY.
FT DISULFID 98 107 BY SIMILARITY.
FT DISULFID 111 267 BY SIMILARITY.
FT DISULFID 254 258 BY SIMILARITY.
FT DISULFID 272 427 BY SIMILARITY.
FT SITE 87 89 CELL ATTACHMENT SITE (POTENTIAL).
FT CARBOHYD 61 61 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 230 230 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 280 280 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 390 390 N-LINKED (GLCNAC...) (POTENTIAL).
SQ SEQUENCE 427 AA; 47413 MW; EAS8631F3EB6047 CRC64;

Query Match 5.3%; Score 663; DB 1; Length 427;
Best Local Similarity 40.3%; Pred. No. 3.6e-27;
Matches 142; Conservative 61; Mismatches 123; Indels 26; Gaps 6;

QY 2018 ECLIGELHAGNSTFLFLVYSNK-----COTPLGNASGHIRDFOITASQY-- 2062
Db 78 KCLVTEDTQRG--DIFTEYICQCPVGYSGIHCELCGSKGLEGAIAADSIASSVVMG 135
QY 2063 ----GQWAPKARLHYSGNSINAW--STKEPSPWIKVDLLAPMIHGIKTOGAKQFPSSLY 2116
Db 136 FMGLQWGPFLARLYRTGTVNAWTASSYDSKFWIQVDFLRKMRVSGVMTQGSASRAGRY 195
QY 2117 ISQFIIMYSLGKWKQVTRGNSGTGLMVFFGNVDSSGKHNIFNPPIIARYIRLPHTHYS 2176
Db 196 LKTFKVAVSLDGRPEFTQDESGTGKDFMGNDNNSKINNFNTLEAQLRLPVSCH 255
QY 2177 INSTLRMLMGDLNSCSMPGMSKAISSDAQITASSYFT--NMFA-TWSPSKARLHLQG 2233
Db 256 RCTLRFFELGCELHGCSEPLGLKNTIPDSQITASSYKTNWLRAGWYPHLGLDNGQ 315
QY 2234 RSNAPRPQVNPKEWLQVDFQTKMVTGVTGQVKSILTSYVKEFLISSODGHQWILF 2293
Db 316 KINAWTQASNAKELQVDFLQTKKVTGIIQTGARDFGHIQYVASKYVAHSDGQWTVY 375
QY 2294 FONGKVKVYFQGNQDEFTPWANSLDPLTRYLRIHPQSVWHQIALRMEVLGC 2345
Db 376 BEQGTSKVFGQVLDNNSKKNIFEPFVARVVRVLPVLSWHNRITRLBLGC 427

RESULT 11
MFGM_MOUSE
ID MFGM_MOUSE STANDARD; PRT; 463 AA.
AC P21956; P97800;
DT 01-AUG-1991 (Rel. 19, Created)
DT 01-AUG-1991 (Rel. 19, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Lactadherin precursor (Milk fat globule-EGF factor 8) (MFG-B8) (MFGM)
DE (Sperm surface protein Sp47) (MP47).
GN MFG-B8.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.; AND SEQUENCE OF 23-35.
RC TISSUE=Mammary gland;
RX MEDLINE=91045008; PubMed=2122462;
RA Stubbs J.D., Lekutis C., Sirger K.L., Bui A., Yuzuki D.,
RA Srinivasan J., Parry G.;
RT "cDNA cloning of a mouse mammary epithelial cell surface protein
reveals the existence of epidermal growth factor-like domains linked
to factor VIII-like sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 87:8417-8421(1990).
RN [2]
RP SEQUENCE OF 23-463 FROM N.A.

```



```

DR Pfam; PF00754; F5_F8_type_C; 2.
DR SMART; SM00181; EGF; 2.
DR SMART; SM00231; FA58C; 2.
DR PROSITE; PS00022; EGF_1; 2.
DR PROSITE; PS01186; EGF_2; 2.
DR PROSITE; PS00026; EGF_3; 2.
DR PROSITE; PS01285; FA58C_1; 2.
DR PROSITE; PS01286; FA58C_2; 2.
DR PROSITE; PS00022; FA58C_3; 2.
KW Glycoprotein; Repeat; EGF-like domain.
FT DOMAIN 2 41 EGF-like 1.
FT DOMAIN 44 88 EGF-like 2.
FT DOMAIN 91 247 F5/F8 TYPE C 1.
FT DOMAIN 252 409 F5/F8 TYPE C 2.
FT SITE 67 69 CELL ATTACHMENT SITE (POTENTIAL).
FT DISULFID 6 17 BY SIMILARITY.
FT DISULFID 11 29 BY SIMILARITY.
FT DISULFID 31 40 BY SIMILARITY.
FT DISULFID 91 247 BY SIMILARITY.
FT DISULFID 234 238 BY SIMILARITY.
FT DISULFID 252 409 BY SIMILARITY.
FT CARBOHYD 41 41 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 372 372 N-LINKED (GLCNAC...) (POTENTIAL).
SQ SEQUENCE 409 AA; 45725 MW; E0C07AF80C29927A CRC64;

Query Match 5.2%; Score 650; DB 1; Length 409;
Best Local Similarity 39.3%; Pred. No. 1.6e-26;
Matches 139; Conservative 62; Mismatches 125; Indels 28; Gaps 6;

QY 2018 ECLIGEHLAGHSTPLVYSNK-----CQPLGASGHIRDPQTASQY-- 2062
DB 58 EEVEIDDAHQ--DVFTYICPHGYTGHCIIICNAPLGMETGAIDFQISASSMHG 115
QY 2063 ---GOWAPKLARLVSGSINAW--STKEPFWIKVDLLAPMIIGIKTGARQKPSIY 2116
DB 116 FMGLQWAPELARLRHAGLVNATSNYDNRNFIQVNLRRMRVTGVVTOGASRAGSAY 175
QY 2117 ISQFIIMSLDGKKQYTRGNSGTGTLVFFGNVDSGIGKHENFPDIIARYIRLRPHYS 2176
DB 176 MKTEKVAYSTDKRKQFIQGAESGDKIPMGNLDSGLKVNLFVPLEVQVYVRLPIICH 235
QY 2177 IRSTIRMEIMGDLSCSMPLGMSKAISDAQITASSYTN---MFATWSPSKARLHOG 2233
DB 236 RCTIRFELFGLGELSCABFLGKNTIPNKQITASSFYRTWGLSAFWSYPTFYRLDNG 295
QY 2234 RSNARPOVANKPEMLQVDFQTKMTKVTGVTTOGVKSLTSMYKBLISSODGHWTLF 2293
DB 296 KFNATQNSASEWLQIDLGQRVVTGIIITQARDFGHIQVAAYKVAYSDGVSWTEY 355
QY 2294 FONGKV--KVPQGNODSFTPVVNSLDPPLRLYRIHPSQWHQIALRMEVLGC 2345
DB 356 RDQGALEGIKFFGNLNNSHKKNMPTETPLTRFVRLPVAWENRITLRVLLGC 409

RESULT 13
MFGN_BOVIN
ID MFGM_BOVIN STANDARD; PRT; 427 AA.
AC Q95114; P79344; Q27959;
DT 01-NOV-1997 (rel. 35, Created)
DT 01-NOV-1997 (rel. 35, Last sequence update)
DT 15-MAR-2004 (rel. 43, Last annotation update)
DE Lactadherin precursor (Milk fat globule-EGF factor 8) (MFG-EG)
DE (MGPS7/53) (PAS-6/PAS-7 glycoprotein) (MFGM) (Sperm surface protein
DE SP47) (BP47) (Components 15/16).
GN MFG8.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A., PARTIAL SEQUENCE, AND CARBOHYDRATE-LINKAGE SITES.
RC STRAIN=Holstein; TISSUE=Mammary gland;

```

```

RX MEDLINE=970C8954; PubMed=8656064;
RA Hvarregaard J., Andersen M.H., Berglund L., Rasmussen J.T.,
RA Petersen T.E.;
RT "Characterization of glycoprotein PAS-6/7 from membranes of bovine
RT milk fat globules.";
RL Eur. J. Biochem. 240:628-636(1996).
RN [2]
RP SEQUENCE OF 18-427 FROM N.A.
RC TISSUE=Mammary gland;
RX MEDLINE=96125736; PubMed=8541316;
RA Aoki N., Kishi M., Taniguchi Y., Adachi T., Nakamura R.,
RA Matsuda T.;
RT "Molecular cloning of glycoprotein antigens MGPS7/53 recognized by
RT monoclonal antibodies raised against bovine milk fat globule
RT membrane.";
RL Biochim. Biophys. Acta 1245:385-391(1995).
RN [3]
RP SEQUENCE OF 19-427 FROM N.A.
RC TISSUE=Testis;
RA Ensslin M.A.;
RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE OF 140-146; 174-187; 233-246 AND 422-427.
RC TISSUE=Milk;
RX MEDLINE=93250576; PubMed=8485470;
RA Mather I.H., Banghart L.R., Lane W.S.;
RT "The major fat-globule membrane proteins, bovine components 15/16 and
RT guinea-pig GP 55, are homologous to MGF-E8, a murine glycoprotein
RT containing epidermal growth factor-like and factor V/VIII-like
RT sequences.";
RL Biochem. Mol. Biol. Int. 29:545-554(1993).
CC -!- FUNCTION: Probably associates with phospholipids on the surface of
CC mammary epithelial cells and milk fat globules. Zona pellucida-
CC binding protein.
CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Name=Long;
CC IsoId=Q95114-1; Sequence=Displayed;
CC Name=Short;
CC IsoId=Q95114-2; Sequence=VSP_001398;
CC -!- TISSUE SPECIFICITY: Milk and spermatozoan.
CC -!- PTM: THE 2 O-LINKED GLYCANS CONSTITUTE OF GAL, GLCNAC AND FUC, WITH
CC PROBABLY FUC AS REDUCING TERMINAL SUGAR.
CC -!- SIMILARITY: Contains 2 EGF-like domains.
CC -!- SIMILARITY: Contains 2 F5/8 type C domains.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X91895; CAA62997.1; -
DR EMBL; S80643; AAB35894.2; -
DR EMBL; Y11719; CAA72406.1; -
DR PIR; S74211; S74211.
DR HSSP; S00740; IIXA.
DR InterPro; IPR006209; EGF like.
DR InterPro; IPR000421; FA58C.
DR InterPro; IPR008979; Gal_bind_like.
DR InterPro; IPR006210; IEGF.
DR Pfam; PF00008; EGF; 2.
DR Pfam; PF00754; F5_F8_type_C; 2.
DR SMART; SM00181; EGF; 2.
DR SMART; SM00231; FA58C; 2.
DR PROSITE; PS00022; EGF_1; 2.
DR PROSITE; PS01186; EGF_2; 2.
DR PROSITE; PS00026; EGF_3; 2.
DR PROSITE; PS01285; FA58C_1; 2.
DR PROSITE; PS01286; FA58C_2; 2.
DR PROSITE; PS00022; FA58C_3; 2.

```

Signal: Glycoprotein; Milk; Repeat; EGF-like domain;
KW Alternative splicing.
FT SIGNAL 1 18
FT CHAIN 19 427 LACTADHERIN.
FT DOMAIN 20 59 EGF-LIKE 1.
FT DOMAIN 62 106 EGF-LIKE 2.
FT DOMAIN 129 265 FS/8 TYPE C 1.
FT DOMAIN 270 427 FS/8 TYPE C 2.
FT SITE 65 87 CELL ATTACHMENT SITE (POTENTIAL).
FT DISULFID 24 35 BY SIMILARITY.
FT DISULFID 29 47 BY SIMILARITY.
FT DISULFID 49 58 BY SIMILARITY.
FT DISULFID 66 77 BY SIMILARITY.
FT DISULFID 71 94 BY SIMILARITY.
FT DISULFID 96 105 BY SIMILARITY.
FT DISULFID 109 265 BY SIMILARITY.
FT DISULFID 252 266 BY SIMILARITY.
FT DISULFID 270 427 BY SIMILARITY.
FT CARBOHYD 27 27 O-LINKED (FUC. . .) (IN PAS-6).
FT CARBOHYD 34 34 O-LINKED (FUC. . .) (IN PAS-7).
FT CARBOHYD 59 59 N-LINKED (GLNAC. . .) (HYBRID)
FT CARBOHYD 227 227 N-LINKED (GLNAC. . .) (HIGH MANNOSE)
FT VARSPLIC 169 221 Missing (in isoform Short).
FT CONFLICT 19 19 A > P (IN REF. 1).
FT CONFLICT 28 28 L > O (IN REF. 1).
FT SEQUENCE 427 AA; 47411 MW; 4CBEE3A1DC4E24 CRC64;

Query Match 5.1%; Score 635; DB 1; Length 427;
Best Local Similarity 37.9%; Pred. No. 1e-25;
Matches 135; Conservative 66; Mismatches 123; Indels 32; Gaps 7;

QY 2018 ECLIGEHLAGSTFLVYSNK-----CQPLGASGHIRDPOITASQY-- 2062
Db 76 ECQVTDSDSHG--DVFIQVTKCPGLGVGHCHBTCTSPGLMGTGAIDSGISASMHIG 133

QY 2063 ----GOWAPKLARLHYSGINAWST---KEPFSWIKVDLAPMIHTGIKTOGARQKFS 2114
Db 134 FMCGLQWAPELARLHQTGINAWSTGYNDKPP--WQVNLMRKMWTVGVVTOQASRAGSA 191

QY 2115 LIISQFIINSLDGKQWYRGNSTOTLVFFQNVDSGKHNIFNPPIARIYIRLHPHT 2174
Db 192 EYLKTEKVAYSTDGRQFQTOVAGRSCKIFIGNVANSGLKINLPTDPLETVQLVLPVPII 251

QY 2175 YSIRSLRMELMGCDLSCSMPLGMSKALSDAQITASSYFTN---MPATWSPKARLHL 2231
Db 252 CHRGCTLRPELLGCELANGTEPEJGLKNDTIPNKQITASSYKWTWGLSAFSPFPYARLDN 311

QY 2232 QGRSNARFRQVNNPKWLQVDFQKTMKVTGVTQTQGVKSLTSMYKFEFVLISSQDGHQWT 2291
Db 312 QGKFNAWTATQNSASEMLQIDJGSQKEVTGIITQGARDFGHIOYVAAYRVAGDDGVWT 371

QY 2292 LFCQNG--KVKVQGNQDSFTFVNSLDPLLPRIYRIHPQSWHQAIRMEVLGC 2345
Db 372 EYKDPGASKEIPFGNNDNNSHRKNIFEPFQARFVRIQPVAMHNRIITLRVELLGC 427

RESULT 14
MFGM_HUMAN
ID MFGM_HUMAN STANDARD; PRT; 387 AA.
AC Q08431;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Lactadherin precursor (Milk fat globule-EGF factor 8) (MFG-8) (HMFG)
DE (Breast epithelial antigen BA46) (MFGM) [Contains: Medin].
GN MFG8.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OX NCBI_TaxID=9606;

[1] SEQUENCE FROM N.A.
RP TISSUE=Breast, and Breast carcinoma;
RC MEDLINE=96213908; PubMed=8639264;
RX Couto J.R., Taylor M.R., Godwin S.G., Ceriani R.L., Peterson J.A.;
RA "Cloning and sequence analysis of human breast epithelial antigen
RT BA46 reveals an RGD cell adhesion sequence presented on an epidermal
RT growth factor-like domain";
RL DNA Cell Biol. 15:281-286(1996).
[2]
RN SEQUENCE OF 170-387 FROM N.A.
RP TISSUE=Mammary Gland;
RC MEDLINE=91371351; PubMed=1909932;
RX Larocca D., Peterson J.A., Ureia R., Kuniyoshi J., Bistrain A.M.,
RA Ceriani R.L.;
RT "A Mr 46,000 human milk fat globule protein that is highly expressed
RL in human breast tumors contains factor VIII-like domains.";
RN Cancer Res. 51:4994-4998(1991).
[3]
RP PARTIAL SEQUENCE, AND CHARACTERIZATION.
RC TISSUE=Milk;
RX MEDLINE=98194924; PubMed=9535276;
RA Giuffrida M.G., Cavalletto M., Giunta C., Conti A.,
RA Godovac-Zimmermann J.;
RT "Isolation and characterization of full and truncated forms of human
RL breast carcinoma protein BA46 from human milk fat globule membranes.";
RN J. Protein Chem. 17:143-148(1998).
[4]
RP SEQUENCE OF 268-317, AND IDENTIFICATION OF MEDIN.
RX MEDLINE=99342076; PubMed=10411933;
RA Haeggqvist B., Naeslund J., Sletten K., Westermark G.T., Mucchiano G.,
RA Tjernberg L.O., Nordstedt C., Engstrom U., Westermark P.;
RT "Medin: an integral fragment of aortic smooth muscle cell-produced
RL lactadherin forms the most common human amyloid.";
RN Proc. Natl. Acad. Sci. U.S.A. 96:8669-8674(1999).
[5]
RP CHARACTERIZATION.
RX MEDLINE=97405885; PubMed=9260929;
RA Taylor M.R., Couto J.R., Scallan C.D., Ceriani R.L., Peterson J.A.;
RT "Lactadherin (formerly BA46), a membrane-associated glycoprotein
RC expressed in human milk and breast carcinomas, promotes Arg-Gly-Asp
RT (RGD)-dependent cell adhesion";
RL DNA Cell Biol. 16:861-869(1997).
CC -!- FUNCTION: May be involved in phospholipid binding. Binds
CC specifically to rotavirus and inhibits its replication.
CC -!- FUNCTION: Medin is the main constituent of aortic medial amyloid.
CC -!- SUBCELLULAR LOCATION: Peripheral membrane protein.
CC -!- TISSUE SPECIFICITY: Mammary epithelial cell surfaces and aortic
CC media. Overexpressed in several carcinomas.
CC -!- PTM: MEDIN HAS A RAGED N-TERMINUS WITH MINOR SPECIES STARTING AT
CC AMINO ACID 264 AND 273.
CC -!- SIMILARITY: Contains 1 EGF-like domain.
CC -!- SIMILARITY: Contains 2 FS/8 type C domains.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/)
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U58516; AAC50549.1; -
CC EMBL; S56151; AAS19771.1; -
CC FIR; A47285; A47285.
CC HSSP; P08709; 1BP9.
CC Genew; HGNC:7036; MFG8.
CC MIM; 602281; -
CC GO; GO:0007155; P-cell adhesion; TAS.
CC InterPro; IPR001438; EGF_II.
CC InterPro; IPR006209; EGF_like.
CC InterPro; IPR000421; FA56C.
CC InterPro; IPR008979; Gal_bind_like.


```
DR InterPro; IPR006210; IEGF.
DR Pfam; PF00008; EGF; 1.
DR Pfam; PF00734; F5/F8_type_C; 2.
DR PRINTS; PR00010; EGFELOOD.
DR SMART; SM00181; EGF; 1.
DR SMART; SM00231; FAS8C; 2.
DR PROSITE; PS00022; EGF_1; 1.
DR PROSITE; PS01186; EGF_2; 1.
DR PROSITE; PS00026; EGF_3; 1.
DR PROSITE; PS01285; FAS8C_1; 2.
DR PROSITE; PS01286; FAS8C_2; 2.
DR PROSITE; PS00022; FAS8C_3; 2.
DR Signal; Glycoprotein; Milk; Repeat; EGF-like domain; Amyloid.
FT SIGNAL 1 23
FT CHAIN 24 387
FT CHAIN 202 387
FT CHAIN 285 317
FT DOMAIN 24 67
FT DOMAIN 70 225
FT DOMAIN 230 387
FT SITE 46 48
FT SITE 27 38
FT DISULFID 32 55
FT DISULFID 57 66
FT DISULFID 70 225
FT DISULFID 212 216
FT DISULFID 230 387
FT CARBOHYD 238 238
FT CARBOHYD 325 325
FT CARBOHYD 329 329
FT CARBOHYD 350 350
SQ SEQUENCE 387 AA; 43123 MW; 2EE6571DEC83782D CRC64;

Query Match 4.7%; Score 588; DB 1; Length 387;
Best Local Similarity 37.3%; Pred. No. 2.3e-23;
Matches 132; Conservative 69; Mismatches 125; Indels 28; Gaps 9;

QY 2066 EMLPGKAGIRVECLIGELHAGMSTFLVYSNKCCTPLCMASGHIRDFQITASG----- 2060
DQ 48 DYFSP-----YTCCLKG---YAGNHC-----ETKVEPLGMEGNTANSQIASSVRYTF 95
QY 2061 -QYGWAPKALRLHVGSGINAW--SKYEFMSIKVDLLAPZMIHGTQCARQKPSLSYI 2117
DQ 96 LGLQHWPELALINRAGMVAWTFSSNDNDNPVQVNLRLRWVTVGTQAGRLAGHEYL 155
QY 2118 SFLIMYSLDGKKWQYRCNSTGLMVFSGNVDSSGIKENINPPIIAYIRLHPHYSI 2177
DQ 156 KAFKVAISLUNGHEFD-FIHDVKKHKEFVGNMKNNAVHVNLFETPVEAQYVRLYPTSCHT 214
QY 2178 RSTLRMLMGCLINSCSMELGMSKASDAQITASSYF----TNMFATWSPGKARLHLQ 2233
DQ 215 ACTLRFELGCELANGCANPLGKNNIPDKQITASSSYKTWGLHLP-SNPSYARLDKQ 273
QY 2234 RNANRPQVNNKEMLOVDFQKTMKVTVGTQVKKLSMTVMYKVEPLISSQDGHQWTLF 2293
DQ 274 NFNAVAGSYGNDQLQVLDLGGSKVETGIIQTQARNFGSVQVFVASKYVAYSNDSANWTEY 333
QY 2294 F--ONGKVKVFCNGQDSFPPVNSLDPBLTLRLTHPQSWHQIALRMEVLGC 2345
DQ 334 QDPRTGSSKIFPGWNDNHHKGNLFTPLIRVIRLPVAVHNRIRARLELLGC 387

RESULT 15
ID NRP2_HUMAN STANDARD; PRT; 931 AA.
AC Q60462; O14820; O14821;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Neurophilin-2 precursor (Vascular endothelial cell growth factor 165
DE receptor 2)
GN NRP2 OR VEGF165R2.
OS Homo sapiens (human).
```

DR SMART; SM00231; FA58C; 2.
DR SMART; SM00137; MAX; 1.
DR PROSITE; PS01140; CUB; 2.
DR PROSITE; PS01285; FA58C_1; 2.
DR PROSITE; PS01286; FA58C_2; 2.
DR PROSITE; PS00222; FA58C_3; 2.
DR PROSITE; PS00060; MAM 2; 1.
DR Transmembrane; Glycoprotein; Neurone; Signal; Repeat; Receptor;
KW Alternative splicing
FT SIGNAL 1 20 OR 22 (POTENTIAL).
FT CHAIN 21 931 NEUROFILIN-2.
FT DOMAIN 21 864 EX-TRACELLULAR (POTENTIAL).
FT TRANSMEM 865 889 POTENTIAL.
FT DOMAIN 890 931 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 28 142 CUB 1.
FT DOMAIN 149 267 CUB 2.
FT DOMAIN 277 427 FS/8 TYPE C 1.
FT DOMAIN 434 592 FS/8 TYPE C 2.
FT DOMAIN 642 802 MAM.
FT DOMAIN 671 674 POLY-SER.
FT DISULFID 28 55 BY SIMILARITY.
FT DISULFID 83 105 BY SIMILARITY.
FT DISULFID 149 175 BY SIMILARITY.
FT DISULFID 208 230 BY SIMILARITY.
FT DISULFID 277 427 BY SIMILARITY.
FT DISULFID 434 592 BY SIMILARITY.
FT CARBOHYD 152 152 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 157 157 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 629 629 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 839 839 N-LINKED (GLCNAC. .) (POTENTIAL).
FT VARSPPLIC 809 813 Missing (in isoform A17).
FT VARSPPLIC 809 830 /FTid=VSP_004341.
FT CONFLICT 602 602 Missing (in isoform A0).
FT CONFLICT 602 602 E->K (IN REP. 1).
FT SSSEQUENCE 931 AA; 104830 MW; 270CBAB69A0A797C CRC64;

Search completed: April 13, 2004, 14:05:08
Job time : 39 secs

Query Match 3.8%; Score 469.5; DB 1; Length 931;
Best Local Similarity 28.3%; Pred. No. 1.2e-16;
Matches 166; Conservative 87; Mismatches 185; Indels 145; Gaps 32;
QY 1847 DEFDCX-ANAYPSDVLDLEK-----DVHSLIGPLLVCHTNL-----NPAHGRQVTOEFA 1896
DB 79 EKHDCKYDFTEIRDGSEADLLGKHCNIAPPTIISSGMLYIKFTSDYARQGA--GFS 136
QY 1897 LFTTFIDETKSWYPTENMERNCRAPCNIOMEDPTFKENYRPHAINGYIMDTLPGLVMAQD 1956
DB 137 LRYEIP-KTGS-----EDCSKNFTSP-NGTIESGFPPEKYP----- 170
QY 1957 QRIRWYLLSGSNENIHSIHPGSHVFTVRKKEYYKALYNLYPGVFTVEMLPSKAG-- 2013
DB 171 -----ENLDCFTTILAKPKMEIILQFL--IFD-LEHDPQLQVGEGD 207
QY 2014 -----IM-----RVECLIGH-----LHAGMST-----LFLV 2035
DB 208 CKYDMLDIWDGPIHVGVLGKYCOTKTPSELRSSTGILSTFFHDMAVAKDGFSAARYLV 267
QY 2036 YSN-----KQOTPLGMASGHIRDQITASQY--GOWAPKLARLHYSGSINAW-----ST 2083
DB 268 HQELENFQCNVPLMGESGRIANEQISASTYSDGRWTPQOSRLH--GDDNGWTFNLDN 325
QY 2084 KEPPSWIKVLLAPMIHGIKTOQA--RQFPSSLYISQFIMYSLDGKKWQTVRGNSTGT 2141
DB 326 KE---YLQVDLRLFTLMTATATQAGISRETQNGYVXSYKLEVSTNGEDWVYRHGKNH- 381
QY 2142 LMVFFGVNDSSGIGENIFNPPIARYIRLPHETHYSIRSTIRMEMLMGCDLNS--CSMP2LGM 2199
DB 382 -KVFQANDDADEVVLLKHLHAPLLTRFVRIRPQTHSGIALRLLELFGCRVTDAPCSNMLGM 440
QY 2200 ESKAISDAQITASSYTNMPATMSPSKARLHLOGRSNAW---RQVNNP?KEWLOVDFQKT 2256
DB 441 LSGLIADSGQISASS--TQEY-LMSPSARL-VSSRS-GWFPRI?QAQPGGEWLQVDLGT 495

GenCore version 5.1.16
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: Apr-11 13, 2004, 14:01:27 ; Search time 68.5 Seconds
(without alignments)
10828.953 Million cell updates/sec

Title: NP000123-328

Perfect score: 12416

Sequence: 1 MOELSTCFILCLRLFCFSA.....WVHQIALRMEVLGCEAQDLY 2351

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- 1: sp_archaea.*
- 2: sp_bacteria.*
- 3: sp_fungi.*
- 4: sp_human.*
- 5: sp_invertebrate.*
- 6: sp_mammal.*
- 7: sp_mhc.*
- 8: sp_organelle.*
- 9: sp_phase.*
- 10: sp_plant.*
- 11: sp_rodent.*
- 12: sp_virus.*
- 13: sp_vertebrate.*
- 14: sp_unclassified.*
- 15: sp_rvirus.*
- 16: sp_bacteriap.*
- 17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	9485	76.4	2343	6 O18806	O18806 canis famil
2	9439	76.0	2343	6 O62730	O62730 canis famil
3	6283.5	50.6	2258	11 Q7TN96	Q7TN96 rattus norv
4	4043.5	24.5	1639	13 Q804W6	Q804W6 fugu rubrip
5	2825.5	22.8	2119	13 Q90X47	Q90X47 brachydanio
6	2774.5	22.3	2183	11 Q88783	Q88783 mus musculu
7	2698	21.6	1802	13 Q804W5	Q804W5 fugu rubrip
8	2511.5	20.2	1377	13 Q804X3	Q804X3 gallus gall
9	2340	18.8	1460	13 Q7SZN0	Q7SZN0 pseudonaja
10	1999.5	16.2	2102	11 Q7TFK2	Q7TFK2 rattus norv
11	1595.5	12.9	355	11 Q8BQ43	Q8BQ43 mus musculu
12	1453	11.7	745	13 Q804X4	Q804X4 gallus gall
13	1389	11.2	1156	11 Q80Y80	Q80Y80 mus musculu
14	1388	11.2	1157	11 Q920Z4	Q920Z4 mus musculu
15	1374	11.1	1157	11 Q920H8	Q920H8 rattus norv
16	1344	10.8	1158	4 Q9BQS7	Q9BQS7 homo sapien

17	1339	10.8	1158	4 Q9C058	Q9C058 homo sapien
18	1328	10.7	1087	13 Q7ZU12	Q7ZU12 brachydanio
19	1307	10.5	1104	4 Q75180	Q75180 homo sapien
20	1298	10.5	1084	11 Q9JL97	Q9JL97 rattus norv
21	1227.5	9.9	1048	6 Q9XT27	Q9XT27 ovis aries
22	1123.5	9.0	847	11 Q8C4S2	Q8C4S2 mus musculu
23	1104	8.9	216	4 Q14286	Q14286 homo sapien
24	932	7.5	626	13 Q90ZT2	Q90ZT2 brachydanio
25	911	7.3	407	13 Q8AY50	Q8AY50 brachydanio
26	743.5	6.0	1142	10 Q8LL91	Q8LL91 chlamydomon
27	700	5.6	503	11 Q8BV37	Q8BV37 mus musculu
28	664	5.3	463	11 Q9R1X9	Q9R1X9 mus musculu
29	662.5	5.3	480	4 Q8N610	Q8N610 homo sapien
30	662.5	5.3	480	4 Q43854	Q43854 homo sapien
31	661	5.3	426	11 Q9WTS3	Q9WTS3 mus musculu
32	655.5	5.3	470	11 Q8C4U8	Q8C4U8 mus musculu
33	655.5	5.3	480	11 Q8CBF7	Q8CBF7 mus musculu
34	655.5	5.3	480	11 Q35474	Q35474 mus musculu
35	555	4.5	312	4 Q7Z3D2	Q7Z3D2 homo sapien
36	514.5	4.1	363	6 Q77718	Q77718 equus cabal
37	498	3.9	335	4 Q9BTL9	Q9BTL9 homo sapien
38	469.5	3.8	901	4 Q9H2E4	Q9H2E4 homo sapien
39	469.5	3.8	901	4 Q9H2D5	Q9H2D5 homo sapien
40	469.5	3.8	906	4 Q9H2D4	Q9H2D4 homo sapien
41	469.5	3.8	906	4 Q9H2E3	Q9H2E3 homo sapien
42	469.5	3.8	931	4 Q7Z3T9	Q7Z3T9 homo sapien
43	466.5	3.8	384	11 Q8C8K0	Q8C8K0 mus musculu
44	462.5	3.7	926	11 Q8QZT7	Q8QZT7 mus musculu
45	450.5	3.6	921	11 Q9QX38	Q9QX38 rattus norv

ALIGNMENTS

RESULT 1

O18806 PRELIMINARY; PRT; 2343 AA.
 AC O18806;
 DT 01-JAN-1998 (TREMREL. 05, Created)
 DT 01-JAN-1998 (TREMREL. 05, Last sequence update)
 DT 01-OCT-2003 (TREMREL. 25, Last annotation update)
 DE Factor VIII.
 GN F8.
 OS Canis familiaris (Dog).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
 OX NCBI_TaxID=9615;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Liver;
 RA Cameron C., Notley C., Hoyle S., McGlynn L., Hough C., Kamisue S.,
 RA "The canine factor VIII cDNA and 5' flanking sequence."
 RT Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
 RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
 CC -!- SIMILARITY: CONTAINS 2 FS/8 TYPE C DOMAINS.
 DR EMBL; AF016234; AAB87412.1; .
 DR HSSP; P0C451; 1CFG.
 DR GO; GO:0005507; P: copper ion binding; IEA.
 DR GO; GO:0007155; P: cell adhesion; IEA.
 DR InterPro; IPR001117; Cu-oxidase.
 DR InterPro; IPR008972; Cupredoxin.
 DR InterPro; IPR000421; Fx58 C.
 DR InterPro; IPR008979; Gal Bind_like.
 DR Pfam; PF003194; Cu-oxidase; 3.
 DR Pfam; PF00754; F5_F8_type_C; 2.
 DR SMART; SM00231; FAS8C; 2.
 DR PROSITE; PS01285; FAS8C_1; 2.
 DR PROSITE; PS01286; FAS8C_2; 2.
 DR PROSITE; PS00022; FAS8C_3; 2.
 DR PROSITE; PS00079; MULTICOPPER_OXIDASE1; 3.
 SQ SEQUENCE 2343 AA; 255629 XM; A854FAE571C3B399 CRC64;

Query Match

76.4%; Score 9485; DB 6; Length 2343;

Best Local Similarity 77.1%; Pred. No. 0;
Matches 1816; Conservative 196; Mismatches 320; Indels 24; Gaps 12;

QY 1 MQIELSTCFELCLLRFCSARRYYLGAVALSWDYMOSD-LGELPVDAREPVRPKSPFF 59
DB 1 MQVELYTCCLLCLLPFSLATRKYYLGAVALSWDYMOSDLSALHATDSFSSRVFGSULP 60

QY 60 NTSVYKTLFVEFTDHLFNIAKPRPMMGLGFTIAQAEVYDVTIVILKNWASHFVSLHA 119
DB 61 TTSVYRKTVFVEFTDHLFNIAKPRPMMGLGFTIAQAEVYDVTIVILKNWASHFVSLHA 120

QY 120 VGVSYWASGAEGAEVDDQTSQREKEDDKVFPQGSHTYVQVILKENGPMASPLCTYSYLS 179
DB 121 VGVSYWASGAEGAEVDDQTSQREKEDDNVFPQGSHTYVQVILKENGPMASPPCLTYSYFS 180

QY 180 HVDLVKOLNSGLIGALLVCBEGSLAKETQTLRKFILLFAVDFDEGKSWHSSTKNSLMODR 239
DB 181 HVDLVKOLNSGLIGALLVCBEGSLAKETQTLRQFVLVFAVDFDEGKSWHSSTKNSLMODR 238

QY 240 DAASARAWPMHTVNGVYNSRSLPGLIGHRKSVYVHWIGMTTPEVHSIFLEGHTFLVRN 299
DB 239 ---AEAQHELHTINGVYNSRSLPGLTVCHKSVYVHWIGMTTPEVHSIFLEGHTFLVRN 294

QY 300 HRQASLEISPIITECTAOTLLMDLGOFLLSCHISSHODHGEAYKYVDSCEBEPOLRMKN 359
DB 295 HRQASLEISPIITECTAOTFLMDLGOFLLPCHIPSHODHGEAYKYVDSCEBEPOLRMKN 354

QY 360 EEAEDYDDDLTDSMDVVRPDDNSPFGIQRSVAKKHPTWVHYIAAEEDWDYAPVL 419
DB 355 ED-KDYDDGLVSDMDVVRPDDNSPFGIQRSVAKKHPTWVHYIAAEEDWDYAPVL 413

QY 420 APDRYSKSOVLNNGPORIGRYKVKRPMAYTDTFTKTRBAIHESGILGPLLYGEGVDT 479
DB 414 TPNDRSKNIYLNNGPORIGRYKVKRPMAYTDTFTKTRBAIHESGILGPLLYGEGVDT 473

QY 480 LLIIFKQASBPYNIYPHGTCTVZPLYSRALPKGVHLKOPPHLPGSLFKYKWTVTYEDG 539
DB 474 LLIIFKQASBPYNIYPHGTCTVZPLYSRALPKGVHLKOPPHLPGSLFKYKWTVTYEDG 533

QY 540 PTKSDPCLTRYYSFVNMRDLASGLIGPLLIYKESVDQRNQIMSKDNVILFSVFD 599
DB 534 PTKSDPCLTRYYSFINERDLASGLIGPLLIYKESVDQRNQIMSKDNVILFSVFD 593

QY 600 ENRSWYLTENLORLPNAPAGVQLSDPPEFQASNIMHSINGVYFVDSLOISVCLHEVAYWYIL 659
DB 594 ENRSWYLTENLORLPNAPAGVQLSDPPEFQASNIMHSINGVYFVDSLOISVCLHEVAYWYIL 653

QY 660 SIGAQTDFLSVFSGYTFKHMVYEDTLTLFPGSGTVMKSMENPGILWILGCHNSDFRNR 719
DB 654 SVGAQTDFLSVFSGYTFKHMVYEDTLTLFPGSGTVMKSMENPGILWILGCHNSDFRNR 713

QY 720 GMTALLKVSSCDKNTGYEDSYEDISAYLLSKNAIEPRSFQNSRHPSTPRQKNATT 779
DB 714 GMTALLKVSSCNRIIDDYEDYEDIEDIPTLLNENNVIKPRSFQNSRHPSTPRQKNATT 773

QY 780 IPENDIEKTDWFAHRTFPMKIQNVSSDLMLLRQSPTHGILSLDQAKYETFSDDP 839
DB 774 TPENDIEKIDLOQSERQOLKAOQVSSDLMLLQNPTRPGLFSLDLREATDR--ADDE 831

QY 840 SPGAIDSNNSISEMTHFRPOLHHSQDMVFTPESGQLRLNEKLGTTAATLKLKLDKVS 899
DB 832 SRGAIERNKGPPEVASLRPELRHSEDEFTPEPELQRLNENIGTNTVELKLDLKLSS 891

QY 900 TSNNLIS--TIPSONLAAGTNTSSLGPPSNVHYDSQOLDTTLFGKSSPSTESSGPLSL 957
DB 892 SSDSLMTSPITPSDKLAATEKTSGLGPPNMSVHFNSHLGTVFNGNNSHLIQSGVPLEL 951

QY 958 SEENNDKLESGLMNSQESWGQVNSTSGRLPKRAKAGPALLTKDNALFKVSTLL 1017
DB 952 SEENDSKLEAPLANTQESLSRENVLMSNRLLFKERIRGPFASLTKDNALFKVNTSSV 1011

QY 1018 KTKNTSNNSATNRKTHIDGSLIENSPSVQNT--LESDFEFKVTPLIHDRMLMOKNAT 1076

DB 1012 KTRAPVNLITNRKTRVAIPTLLIENSTSVQOIMLERNTEFKEVTSLIHNETFMDRNTT 1071

QY 1077 ALRLNHSNKTTSKQMEMVQOKKEGPIPDQONPOMKSFKNLFLPESARWIOETHCKNS 1136

DB 1072 ALGLNHSNKTTSKQMEMVQOKKEGPIPDQONPOMKSFKNLFLPESARWIOETHCKNS 1127

QY 1137 LNSGQSPSPQQLVSLQPEKSVQGNFLSEKNVVGKGETTKDVGLEKEMVFPSSRNFLT 1196

DB 1128 LSSBQSPSPQQLVSLQPEKSVQGNFLSEKNVVGKGETTKDVGLEKEMVFPSSRNFLT 1185

QY 1197 NLNHLNHNTHNQEKLIQEBIEKKEITLIQENVLPQIHTVGTGNPMKNLFLSTRONVE 1256

DB 1186 NLNVOENDTNCCKSPBIEERKEKITQENVALPQAHTMTIGTKNFKLFLSTKQVNA 1245

QY 1257 GSYDGAVAPVLQFRLSNDSTNRKHTAHFKKGBEENLEGLGNQTKQVEXVACTRI 1316

DB 1246 GLREQPTPLQOTRSLNDSPHSEGIHMANFSKIREANLEGLGNQVQWERPPSTIRM 1305

QY 1317 SPNTSQNFYQSRKALKQFRLEBETELEKRIIVDDTSTQWSKNMKHLTPSTLTQIDY 1376

DB 1306 SSNAS-QHVITQRGKSLKQPLSQGEIKFERKVIANDTSTQWSKNMYLAQGTLTQIEY 1364

QY 1377 NEKEGKITOSPLDCLTRSHSIPQANRSELPPIAKVSSPFSIRPIYLVTRVLPODNSSHLP 1436

DB 1365 NEKEGKITOSPLDCLTRSHSIPQANRSELPPIAKVSSPFSIRPIYLVTRVLPODNSSHLP 1424

QY 1437 AA---SYRKKGSGVQESSHFLQAKKNLSLAILTLEMTGDQREVGSIGTSATNSVTYK 1492

DB 1425 ASACNTVFRERTSGVQEGSHFLQAKKNLSLAFVLITGTEGQKPSLQSKSATNQPMYK 1484

QY 1493 KVNTVLPKPDLKPTSGKVELLPKVIYQKDLPTTTSNGSPGHLDLVBESLLOQTEGAI 1552

DB 1485 KLENTVLLQGLSETSKVELLSQHVQDQSDSPFTKTSNDSPGHLDLIMGKIFLQKTOGPV 1544

QY 1553 KWEANRPKVPFLRVATESSAKTPSKLLDPLANDNEHYCTQIPKEBKSSEKSEKTAFAK 1612

DB 1545 KMKNTSPGKVPFLRVATESSAKTPSKLLDPLANDNEHYCTQIPKEBKSSEKSEKTAFAK 1604

QY 1613 KCDTILSLNACENHAIAAINEGONKPEIIVTAKQRTERLCSQNPVPVLRQREITRT 1672

DB 1605 RKDTILPLGFCENNDSTAINEGQDKQREAMAKQGEFGRILCSQNPVPVSKHQREITVT 1664

QY 1673 TLOSDQBEIDYDTISVENMKEDFDYDDBENQSPRSFQKTRHYTIAAVERLWDYGMSS 1732

DB 1665 TLOPEEDKPEYDFTSLEMKREDFDYDGYENQGLRSFQKTRHYTIAAVERLWDYGMSS 1724

QY 1733 SPVLNRAGSGSVPOFKVVFQSPFDGSPQPLYRGELNEHGLLGPYIRASVEDNIMV 1792

DB 1725 SPILLNRAGSGSVPOFKVVFQSPFDGSPQPLYRGELNEHGLLGPYIRASVEDNIMV 1784

QY 1793 TFRNQASRPYSYSSLSIYEEDQORQGAERPKNFVKPNETKTYFWKQVHHMAPTKDFDCK 1852

DB 1785 TFRNQASRPYSYSSLSIYEEDQORQGAERPKNFVKPNETKTYFWKQVHHMAPTKDFDCK 1844

QY 1853 AWAYFSDVLDKDVHSGSLGPLLVCHTNILNPAHQGVTVQEFPALEPTTFDETKSWYFTE 1912

DB 1845 AWAYFSDVLDKDVHSGSLGPLLVCHTNILNPAHQGVTVQEFPALEPTTFDETKSWYFTE 1904

QY 1913 NMERNCRAPCNIOQVEDPTFKENYRFFHAINGYIMDTLPGVMAODQRIWVLLSMGSENI 1972

DB 1905 NLERNCRAPCNVQKEDPTFKENYRFFHAINGYIMDTLPGVMAODQRIWVLLSMGSENI 1964

QY 1973 HSIHFSCHVFTVRKBEYKMAALYNLYPGVETVEMLPKAGIWRVRECLIGEHLHAGMSTL 2032

DB 1965 HSIHFSCHVFTVRKBEYKMAALYNLYPGVETVEMLPKAGIWRVRECLIGEHLHAGMSTL 2024

QY 2033 FLVYSKCCQTPJGASGHIRDFOITASQGVQWAPKLARLHYSGSINAWSTKPPSPWIKV 2092

DB 2025 FLVYSKCCQTPJGASGHIRDFOITASQGVQWAPKLARLHYSGSINAWSTKPPSPWIKV 2084

QY 2093 DLLAPMLIHGKITQGARQKFSLSIYISQFTIMVSLDGKQVYRGNSTGTLYVFFGNVDSS 2152

DB 2085 DLLAPMLIHGKITQGARQKFSLSIYISQFTIMVSLDGKQVYRGNSTGTLYVFFGNVDSS 2144

```

QY 2153 GIKENFNPIIARYIRLHPTHSYRSIRSTRMELMGCDLNSCMSPLGMSKAI SDAQITAS 2212
DB 2145 GIKENFNPIIACYIRLHPTHSYRSIRSTRMELMGCDLNSCMSPLGMSKAI SDAQITAS 2204
QY 2213 SYTNMPTWSPSKARHLQGRSNAMPQVNNPKWLVQDFQTKMKVTVTQGVKSLLT 2272
DB 2205 SYLSSMLATWSPSQARHLQGRSNAMPQVNNPKWLVQDFQTKMKVTVTQGVKSLLI 2264
QY 2273 SMYVKEPLISSQDGHQWTLFFQNGKVKVFGQKODSETPVNSLDPPLLTRYLRHPOSW 2332
DB 2265 SMYVKEPLISSQDGHQWTLFFQNGKVKVFGQKODSETPVNSLDPPLLTRYLRHPOSW 2324
QY 2333 VHOIALREVLGCEAQ 2348
DB 2325 AHHIALREVLGCDTQ 2340

RESULT 2
O62730 PRELIMINARY; PRT; 2343 AA.
ID O62730 AC O62730:
DT 01-AUG-1998 (TrEMBLrel. 07, Created)
DT 01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Factor VII.
OS Canis familiaris (Dog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
OX NCBI_TaxID=9615;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Kidney, and Spleen;
RA Gordy P.W., Bowen R.A.;
RT "Characterization of the canine factor VIII cDNA.";
RL Submitted (FEB-1998) to the EMBL/Genbank/DBJ databases.
CC -1- SIMILARITY: CONTAINS 2 FS/8 TYPE C DOMAINS.
DR EMBL; AF049489; AAC05384.1; -.
DR HSSP; P00451; 1CFG.
DR GO; GO:000507; P: copper ion binding; IEA.
DR GO; GO:0007155; P: cell adhesion; IEA.
DR InterPro; IPR001117; Cu-oxidase.
DR InterPro; IPR008972; Cupredoxin.
DR InterPro; IPR000421; FA58_C.
DR InterPro; IPR008979; Gal bind like.
DR Pfam; PF00394; Cu-oxidase; 3.
DR SMART; SM00231; FA58C_2.
DR PROSITE; PS01285; FA58C_1; 2.
DR PROSITE; PS01286; FA58C_2; 2.
DR PROSITE; PS00222; FA58C_3; 2.
DR PROSITE; PS00079; MULTICOPPER OXIDASE1; 3.
SQ SEQUENCE 2343 AA; 265613 MW; F612D744ADAADD99 CRC64;

Query Match 76.0%; Score 9439; DB 6; Length 2343;
Best Local Similarity 76.7%; Pred. No. 0;
Matches 1808; Conservative 199; Mismatches 325; Indels 24; Gaps 12;

QY 1 MQBELSCFFLLRPFCSATRRYVLGAVELSWDYMOSD-LGELPVDARPPRPVKSPFP 59
DB 1 MQVELYICCFLLLPFLSATRKYVLGAVELSWDYMOSDLSALHADTSFSSRVGSLPL 60
QY 60 NTSVYKKTFLVFETHLNFNTAKPRPPWMLGGFTIQAEVYDVTIVITLKNMASHPVSLHA 119
DB 61 TTSVYKKTFLVFETHLNFNTAKPRPPWMLGGFTIQAEVYDVTIVITLKNMASHPVSLHA 120
QY 120 VGSYVWASGAEAYDDQTSQSEKEDDKVPFGSGSYTYVQVLKNGPMASDPLCLTYSLYS 179
DB 121 VGSYVWASGAEAYDDQTSQSEKEDDKVIPGESHYTYVQVLKNGPMASDPPCLTYSYFS 180
QY 180 HVDLVKDLNSGLIGALLVCKSGSLAKERTQTLQEFVLLFAVFDGKSWHSETKNSLMODR 239
DB 181 HVDLVKDLNSGLIGALLVCKSGSLAKERTQTLQEFVLLFAVFDGKSWHSETNASLTQ-- 238

```

```

QY 240 DAASARAWPMHVTNGVYVNRSLPCLIGCHRSYVWHVIGMGTTPVHVSIFLEGHTFLVRN 299
DB 239 ----AAQAEHLHNTNGVYVNRSLPCLTVCHRSYVWHVIGMGTTPVHVSIFLEGHTFLVGN 294
QY 300 HRQASLEISPIITLTATQTLMDLGQFLHCHISSHQHDGMEAYVKVDSCEEPQLRMKN 359
DB 295 HRQASLEISPIITLTATQTLMDLGQFLHCHISSHQHDGMEAYVKVDSCEEPQLRMKN 354
QY 360 EBAEDYDDDLTDEMDVVRDDDNSSPSFQIRSVAKKHPTWVHYTAAREDDYAPLVL 419
DB 355 ED-KYDDGLYGDMDVSEDDSSSFFQIRSVAKKHPTWVHYTAAREDDYAPSGP 413
QY 420 APDRSYKSOYLNNGPQIRGKYKVRFMAYTDETFKTREAIQHSGLGLPLLYGEGYDT 479
DB 414 TPNDRSHKLYLNNGPQIRGKYKVRFMAYTDETFKTREAIQYESSILGLPLLYGEGYDT 473
QY 480 LLIIFKQASRPNTYPHGTTDVRPLYSRRLPGVKHKLDPFPLPGEIFYKYKWTVTVEDG 539
DB 474 LLIIFKQASRPNTYPHGTTDVRPLYSRRLPGVKHKLDPFPLPGEIFYKYKWTVTVEDG 533
QY 540 PTKSDPRCLTRYSSFWNMRDLASGLIGPLLI CYKESVDORGNOIMSDKENVILFSVFD 599
DB 534 PTKSDPRCLTRYSSFWNMRDLASGLIGPLLI CYKESVDORGNOIMSDKENVILFSVFD 593
QY 600 ENRSWYLTENIQRFPLNPAGVQLEDPEFQASNIHMSINGVYVDSLQLSVCLHEVAYWYL 659
DB 594 ENRSWYLTENIQRFPLNPAGVQLEDPEFQASNIHMSINGVYVDSLQLSVCLHEVAYWYL 653
QY 660 SIGAQTDFLSVFPFGVTFKHQVYEDTLTLPFSGGTVFMVSMENPGLWILGCHNSDPNR 719
DB 654 SVGAQTDFLSVFPFGVTFKHQVYEDTLTLPFSGGTVFMVSMENPGLWILGCHNSDPNR 713
QY 720 GMTALLKVSSCDKNTGVDYEDSYEDISAYLLSKNNAIPEPSFQNSRPHSTROKOFWAT 779
DB 714 GMTALLKVSSCDKNTGVDYEDSYEDISAYLLSKNNAIPEPSFQNSRPHSTROKOFWAT 773
QY 780 IPENDIEKTDTPFAHRTMPKIQNVASSDILLMLRQSPTPHGLSLDLOAKYETTFSDDP 839
DB 774 TPENDIEKTDLOSGERTQLIKAQSVSSDILLMLRQSPTPHGLSLDLOAKYETTFSDDP 831
QY 840 SPGAISNNSLSEMTFRPOLHSGDMVFTPEBSGLQLRLNEKLGTYAATBLKLDKFKVSS 899
DB 832 SRGAIERNKGPPEVASLPELRHSEDRFTPEPELQLRLNENLGTNTVTLKLDLKISS 891
QY 900 TSNNLIS--TIPSDNLAAGTDNTSSIGPPSPMPVHYDSQDQTLFGKSSPLTSSGGPLSL 957
DB 892 SSLSLMSPTIPSDKLAATEKTGSLGPPNWSVHFNGHLGTIVFGNNSHLIGSGVPEL 951
QY 958 SEENDSKLLESLGIMNSQESSWGKNVSSYTESGRLFKGRAGHPALLTKDNALFKVSISSL 1017
DB 952 SEENDSKLLEAFIMNIQESSLRENVLSMESNLFKEERIRGFASLIKDNALFKVNISSV 1011
QY 1018 KTKNTSNNSATNRKTHIDGSLIENSPSVQNI--LESDFPKVYPLIHDRMLMDKNAT 1076
DB 1012 KTRNAPVNLTRKTRVAIPFLIENSSTVQIMLERNTFEFKVTSLIHNETFMDRNT 1071
QY 1077 ALRLNMSNKTTSKKNMVCQKKEGIPDDAQNPDMSPFKMLFLPESABWIORTHGKNS 1136
DB 1072 ALGLNHVSNKTTLSKNVEAHQKEDPVPURBNPDLSSSKIFPLD---WIKTHGKNS 1127
QY 1137 LNSGQSPSPQLVSLGPEKSVGQNFLEKKNVYVGVGGETKDVGLKEMVFPSSRLFLT 1196
DB 1128 LSSEQRSPQLTSLGSEKSVKQNFLESEB-KVVVGDEFTKDTQLQE-IFPNKKSIFA 1185
QY 1197 NLNLNHNTHNCEKKEQIEIEKETLIQENVLPQIHVTGTNRKMNFLILSTRQVE 1256
DB 1186 NLNVQENDTYNCKKSLEIEIEKELTQENVALPQARTWIGTKNFKNLFLSTKQNV 1245
QY 1257 GSYDGAVAPVLOPFRSLNDSTNRKHTAHFYSKGBEENLEGNGNOTKQIVKACTRI 1316
DB 1246 GLEQPYTPILQOTRSLNDSPHSEGHMANFSKIRBEANLEGLNQINQWVFPSTTRM 1305

```

```
QY 1317 SPNTSQNFVTRQSKRALQFRLEPLETELEKRIIVDDTSTQMSKMKHLPSTLTQIDY 1376
Db 1306 SSNAS-CHVITQGRKSLQKPRLSQSGEIKFERKVIANDTSTQMSKKNVLAQGLTLQIEY 1364
QY 1377 NEKEKGAITOSPDLCLTRSHSIFQANRSP:PIAKVSFSPSIRPIYLTRVLFDQNSHLP 1436
Db 1365 NEKEKRALTQSPDSCMRHVTIQMDSALFVAKESASPSVRHTDUTKIPSQCHNSHLP 1424
QY 1437 AA-----SYRKXDSGVQSSHFLOQAKKNLSLAILTEMTGQDQREVGSLSGTSAINSVYIK 1492
Db 1425 ASACNYTFRETSQVQSGSHFQLOEAKKNLSLAFVTLGITGEGQKFSLSGKSAINQPMYK 1484
QY 1493 KVENTVLKPDLPKTSKVELLPKXVHYQKDLFPFTTSNGSPGHLDIVEGSLLOGTGAI 1552
Db 1485 KLENTVLQGLSTSDKVELLSQVHDQEDSFPTKTSNDSPGHLDLMGKIFLQKQGPV 1544
QY 1553 KWNENANPCKVPFURVATESAKTPSKLQPLAWDNHYGTQIPKEENKSOEKSEKTAEX 1612
Db 1545 KANKTNSPGKVPFKWATESSEKIPSKLGLVLANDNHYDTQIPSEEWKSKQSTNTAFK 1604
QY 1613 KDTILSLNACESHAJAAINEGONKPEIEVYMAKQGRTERLCSQNPVPLKRRHQREITRT 1672
Db 1605 RKDTILPLGPCENNDSTAAINEGQDKPQREANMAKQCEPGLCSQNPVSKHHQRE:TVT 1664
QY 1673 TLQSDQEEIDYDDTISVEMKEDPDIDYDENQSPRSFQKTRHYSTAAVERLWDYQMS 1732
Db 1665 TLQDEEDKFEYDDTFSIEMKREDPDIDYDQDQGRSFQKTRHYSTAAVERLWDYQMSR 1724
QY 1733 SPVLRNRAQSGVQPQKVVQFQFTDQSGFTQPLYRGELNEHLGLLGPYIRAEVEDNIV 1792
Db 1725 SPHLNRNRAQSGVQKVVQFQFTDQSGFTQPLYRGELNEHLGLLGPYIRAEVEDNIV 1784
QY 1793 TFRNQAARPSYFSSLSYBDDQQAEPKRNFKYKNETKTYFWKVOHHMAPTKDEEDCK 1852
Db 1785 TFRNQAARPSYFSSLSYBDDQQAEPKRNFKYKNETKTYFWKVOHHMAPTKDEEDCK 1844
QY 1853 AWAYFSVDLEKDVHSLGLLGLLCHTNTLNPAGROVTVQEFALPTTIDETKSWTFTE 1912
Db 1845 AWAYFSVDLEKDVHSLGLLGLLCHTNTLNPAGROVTVQEFALPTTIDETKSWTFTE 1904
QY 1913 NMBRNCRAPCNIQMEDPTFKENYFHAINGYIMDTPLGLVNAQDORLRYLLSGNSNI 1972
Db 1905 NMBRNCRAPCNIQMEDPTFKENYFHAINGYIMDTPLGLVNAQDORLRYLLSGNSNI 1964
QY 1973 HSHFSGHVTVRKKEZYKVALYNLYPGVFTVEMLPKSGAGIWECELI GEHLHAGMSTL 2032
Db 1965 HSHFSGHVTVRKKEZYKVALYNLYPGVFTVEMLPKSGAGIWECELI GEHLHAGMSTL 2024
QY 2033 FLVYSNKCQPLGWSGHIRDFOITASGOXGOWAPKLARLHYSGSINAWSTKBPFSWKV 2092
Db 2025 FLVYSNKCQPLGWSGHIRDFOITASGOXGOWAPKLARLHYSGSINAWSTKBPFSWKV 2084
QY 2093 DLLAPMIHGKQGARQKFSYISQFIIMYSLDGKKWQTYRNGNSTGTGLMVVFGNDSS 2152
Db 2085 DLLAPMIHGKQGARQKFSYISQFIIMYSLDGKKWQTYRNGNSTGTGLMVVFGNDSS 2144
QY 2153 GIKHNIENPILARYBLHPTHYSIRSLRMLMGCDLNCSPMLGSKAISDAQITAS 2212
Db 2145 GIKHNIENPILARYBLHPTHYSIRSLRMLMGCDLNCSPMLGSKAISDAQITAS 2204
QY 2213 SYFTNMEATWSPSKARLHLQGRSNAPVQVNNPKWEI:QVDFQKTMKVTGVTQGVKSLLT 2272
Db 2205 SYLSSMLATWSPSKARLHLQGRSNAPVQVNNPKWEI:QVDFQKTMKVTGVTQGVKSLLT 2264
QY 2273 SMYVKEFLISSQDGHQWTLFFQNGKVKVQGNODSFTPVVNSLDDPPLLTRYLRHQPQSW 2332
Db 2265 SMYVKEFLISSQDGHQWTLFFQNGKVKVQGNODSFTPVVNSLDDPPLLTRYLRHQPQSW 2324
QY 2333 VQIQTALRMEVLGCEAQ 2348
Db 2325 AHHIALRLVLEIGCDTQ 2340
```

```
RESULT 3
QY 07TN96 PRELIMINARY; PET; 2258 AA.
ID 07TN96 AC 07TN96;
AC 07TN96;
DT 01-OCT-2003 (TEMBLrel. 25, Created)
DT 01-OCT-2003 (TEMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TEMBLrel. 25, Last annotation update)
DE Factor VIII.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
CX NCBI_TaxID=101116;
RN [1]
SEQUENCE FROM N.A.
RP STRAIN=Mistar; TISSUE=Liver;
RA Watzka M., Geisen C., Seifried B., Oldenburg J.;
RT "Sequence of the rat factor VIII cDNA.";
RL Submitted (AUG-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY362193; AAQ21580.1; -.
SQ SEQUENCE 2258 AA; 251299 MW; 13AF91C788059B1D CRC64;

Query Match 50.6%; Score 6283.5; DB 11; Length 2258;
Best Local Similarity 53.2%; Pred. NO. 0;
Matches 1272; Conservative 317; Mismatches 629; Indels 171; Gaps 29;

QY 1 MQIELSTCFCLLRFCFSATREYVYLGAVELSDYMQSDL-GELPYDARPPPPVPSFPF 59
Db 1 MRAARGLCFPLSLCALCSNATRYVYLGAVELPDYVSGGASGAQSRSPFPPTAP-- 57
QY 60 NTSVNVYKTLFVETDHLNIAKPRPWWMLGLOPTQAEVYDVTWITLKMASHPVSLHA 119
Db 58 SAHVHTKTVFVYMDRPFHTAXPRPLMWGLLPTITVEHDTWITLKMASHPVSLHA 117
QY 120 VGVSYKASGAYDDQTSQREDEKVPFGSGHTYVQVLKENGPMASDPLCLTYSL 179
Db 118 VGVSWKASGAYDDHSSPAEKDDKVLPGESHTYAWQVLGSGPMASDPLCLTYSL 177
QY 180 HYDLVKDNLGSLGALLVCRGSLAKEKTQLHKFILLFAVDEGKSWHSETKNSLMQDR 239
Db 178 HYDLVKDNLGSLGALLVCKEGLSAERTWMPPEFVILLFAVDEGSRWHAAT--R 230
QY 240 DAASAKAWPMHTVNGVYVNSLPGLIGCHRKSYVHWIYMGTTPEVHSIFLEGHTLVN 299
Db 231 DPASTEAOQTVHNGYTNRTPLGLTGCRTSYVHWMAVGTTPDIESILLEGHTFSVS 290
QY 300 HRAQSLIEIPITELTAQTLIMDLGQFLSCHISSHQHDMGMEAYVKYDSCPPEPOLMK 359
Db 291 HRAQSLIEIPMTLLTAQTLMDLGRFLSCHISSHRHGMGMEAYVQVDCPEAQGGKRD 350
QY 360 EEAEDYDDDLTDSEMDVVRPDDNDSPTQIRSVAKKPKTWVHYIAABEEDWDYAPLV 419
Db 351 EDEEDYDD--LDSEMDVFTWDPDAAP-FVQVRSAAKRPKPTWVHYIAABEGWDYAPAA 407
QY 420 APDRSVKSOYNNGQRTGKRYKVRMAVYDTEFTKTRALQHESGILGLLYGVGD 479
Db 408 TLEDGSRNRYGRGPRIGSKYKRVAYTDGTFTHTREATPREAGLGLLYGVGDS 467
QY 480 LLIIFNQASRPNIYPHGTITDVRPLYSRRLPKGVKHLKDFPILPGEIIFYKWTWVEDG 539
Db 468 LLIVFNKRSRAVNIYPHGIIRDVGAVHAGELPVGKVDLPIRPOETPKYRWLTAE 527
QY 540 PTKSDPRCTRYYSVNNMERDLASGLIPLICYKESVDQRNQIMSKRNVLFSVD 599
Db 528 PARSDARCVTRYASAVDPERDLASGLIPLICCKESVDQRNQIMSDERNVLFSVD 587
QY 600 ENHSWYLTNIGFLFNPAQVLEDEPFOASINMISNGVYFDSLSVCLHEVAWYIL 659
Db 588 ENHSWYISNMRRFLPDEAHVQLQDPEFASINMISNGVYFDSLSVCLHEVAWYIL 647
QY 660 SIGAQTDFLSVFSGYTFXHKMVEYEDTLTLFPFSGTFVMSMENPGLWILGCHNSDFNR 719
Db 648 SVGAQTDFLSVFSGHTFHRVAVYADTLTVFHSYGVTVFMSMDNPGVWVLCNDFRES 707
```

720 GMTALLKVSSCDKNTG3YEDSYEDI5AYLLSKNNALEP3SFQNSRHPSTROKQFNATT 779
708 GMTALLKVSSCDKNTG3YEDSYEDI5AYLLSKNNALEP3SFQNSRHPSTROKQFNATT 763
780 IPENDIEKTDPAHRTTPNPKI QNVSSDLMMLLRQSPHGLSLDLOAKYETSDDP 839
764 -----TPM-----RFLSHPTQGSFWSDSQGDVHDVHRAH 795
840 SPGAIDNNLSLSEMTFRPOLHSGDMVFTPESSLOLRNLKXLTAAATLKLDFKVS 899
796 SPAAHSHNEGPAQAQLPHEGKQKALPSS---WLAARSLVTTWEAKKLLDQVQ- 851
900 TSNLLSTIPSDNLAAGTNTSLGPPSPV-----HYDSQLDTTLTFGKKSSPLTE 950
852 -----VSLGPD-----DRTTAVAPDDVAACKXAGSSGFPDESSPAALGKKMYPR 899
951 SCGPLSLSPENNDSKLLSGLMNSQBSGKNSVSTESGLRFLPKGRAGHALLTKONALF 1010
900 FHGFLSIIEGNRDSNSDSTLMYRLGSPPGDATSWTENRGLGKGRSHRVAFLARGNTLL 959
1011 KVISILLKTKTNNKATNEKTHIDGSLIENSPPVWQNILEDSTEFKVTPLIHDRML 1070
960 -----SDVDEKSHAPGPTSVGNSTAAVQDTILEICSEILEVTPPIHDRIL 1004
1071 MDKNATLRLNEMSKNTTSSKNMEMVQOKKEGPIPPDAQNPDMSPFKMLFELPESARWQR 1130
1005 SDAKATVLRPRTDRTITSRERKDIRHEKODGLVPOQADDTSAFPSEAPLSSTDLWKE 1064
1131 THGKNSLNSQGPSKQVLSGLPEKSVEGQNFISEKKNVVGVEFTKQVGLKZEMVPPSS 1190
1065 ANGDNKAKPQESPPQVLYLWYIKMENQSPSEKKNKVIAGQGFTNTGLEDTVPFRK 1124
1191 RNLFNLNMLNHNENHNOEKKI0EIEKKEKTELIOENVLPLQHTVTGTKNENKMLFLLS 1250
1125 TSVFLTVAKRQBSGRHQB-NIPOAVTEKEAIEBKALPQVHIAAGSKNFRPMFVLG 1183
1251 TRONVGSVDGAVPVLQDPRNSLNTNRKKTATHFSSKGEDE--NLEGLNQTQXIVE 1308
1184 TQONI-NLHEETVVPVHKVWPTRNPTDTRQIPMWHFFKRRKEETNNGSLVKNKTRETVR 1242
1309 KYACTIRISNTSQONFVQSRSEALKQFLPLEETELKRIIVDTSTQKSNMGMHLTP 1368
1243 NY-----PSQKNSVARRRQOAGRIK-----ASARWLPDVARSIQ -277
1369 STLTQIDYNEKEKGAITQSPSLDCLTRSHSIPQANRSPPLPIAKVSSFPSPRIPLRLRVLF 1428
1278 SLLKQIDHRKERRKFIIESRADS-SGTSPTQTNHSPSHVWVMSAPP---PDIIRIIS 1333
1429 QDNSSHLPAASY-----RKXDSGVQSSSHFLQGAKKNLALILEMTGQDREVGSLGTS 1484
1334 RD-SSQVWVSSVAYDFTGSSRIESSPFLNETETDNPSLAVPPRFRVRGRFASPEKV 1392
1485 ATMSVYKKVENTVLKPLDPTKSGVELLPKHVHYQKDLFPNETSGSPGHLDVEGSL 1544
1393 NTHSACDPENNVSEPVSPGBAVTVALPPHVTQEEPLPAGSSCERRGHVLDLFLTS 1452
1545 LQOTEGAIKWNENRPGKVFLEVAETESSAKTPSKLLDPLANDNHYGTQIPKEWKSQEK 1604
1453 LQTRGFEVRSRRRSGG-----DTEGHTENPGKTPSP-----PQMPKQDQSEAG 1499
1605 SP-----EKTAPKKDITLSLNACSNHAIAINQGNQKNETEVTWAKQRTERLCSON 1658
1500 YPKICRWRDGTVLPPPHNLSLGA-----KKKPNLPREARVREVEGGEAPSLVSPE 1549
1659 PPVLKXHQREITR-----TTLQSDQREI--DYDDTISVEMKEDDFDIYBE 1701
1550 PLVLRPRPREASTLLPGGEMREDDGCVTAAYDDDDVMAEYDDAVVTDTEDFDIYBE 1609
1702 DENQSPRSQKTRHYFIAAVERLWDYGMSSSPHVLNRNAQSGVPPQFKKVFQERTDGS 1761
1610 EAGQGRFGQKTRHYFIAAVERLWDYGVVRVSGALGDRAWSDAARFKVFRFRTDGS 1669
1762 FTQPLYRGELNEHGLLGLGYIRAEVEDNINVTRNQASRPYSFYSLISYEEDQROGARP 1821

Db 1670 PTCVHRGELDAHLGULLGYIRAEVEDNINVTRNQASRPYSFYSLVSPED-GBAAP 1728
QY 1822 RNVFVNPENKTYFWKVOHHMASTKDBFCKAWAYFSDVDLEKDVHSLGLIGPLLVCHTNT 1881
Db 1729 RSNFVNPENKTYFWRVRPMAPTDGEFCKAWAYFSDVDLEKDVHSLGLIGPLLVCHAST 1788
QY 1882 LNFAGRQVTVQEFALFFETIDETKSWYFTENMERNCRAPCNTQMEDPFPKENVRHAIN 1941
Db 1789 LHFSHARPLAVQBFALLFAVDETCKSWYCAENLDKCRPRGTPAGDPARWEYRFAVN 1848
QY 1942 GYIMDTLPGLWMAQDQIRWYLLSMGNSNENIHGHSFSGHVFTVRKKEEYKVALYNLYPGV 2001
Db 1849 GYVADALLGLVMAEGRTRWHLLSMGDPGHAQSVHPSAHSVTVRDGGEHTAVCNLYPGV 1908
QY 2002 FETVEMLPKAGIRWRVECLIGEHLHAGMSTLFLVYVYKNCQOTPLGMASGHIRDFQITASGQ 2061
Db 1909 FTTVEMLPKAGIRWRVESLVGEHLRAGMSALFLVYVYKNCQOTPLGMASGHIRDFQITASGQ 1968
QY 2062 YGOWAPKLARLHVSINAWSTKEPESWIKVD-LAPMIHGIKTQAGQKQFSSLYISQFI 2121
Db 1969 HGWTQPLARLHVSINAWSTKEPESWIKVD-LAPMIHGIKTQAGQKQFSSLYISQFI 2028
QY 2122 IMYSLDGKXQWYRGKNTGTLMVFFGNVSSGTHKNIENPPIIARIRLHPTHYVIRSTL 2181
Db 2029 IMYSLDGKXQWYRGKNTGTLMVFFGNVSSGTHKNIENPPIIARIRLHPTHYVIRSTL 2088
QY 2182 RMELMCDLNSCMPGLMESKKAISDAQITASSYFTNMFTWSPSKARLHLOGSNARWQP 2241
Db 2089 RMELMCDLNSCMPGLMESKKAISDAQITASSYFTNMFTWSPSKARLHLOGSNARWQP 2148
QY 2242 VNNKPELQVDFQKMKVTVGTQGVKSLTSMYVKEPLISSODGHQWTLFQNGKVKV 2301
Db 2149 VNDPTQWQVDFQKMKVTVGTQGVKSLTSMYVKEPLISSODGHQWTLFQNGKVKV 2208
QY 2302 FQGNQDSFTPVVNSLDPPLIRYLRHPQSWHQAIALRMEVLGCEAQDL 2350
Db 2209 FQGNRDASTPMVNSLHPFRFYLRIHPQVWEQIALRMEVLGCEAQDL 2257
RESULT 4
Q804W6 PRELIMINARY; PRT; 1639 AA.
AC C804W6;
DT 01-JUN-2003 (TREMELrel. 24, Created)
DT 01-JUN-2003 (TREMELrel. 24, Last sequence update)
DT 01-OCT-2003 (TREMELrel. 25, Last annotation update)
DE Coagulation factor VIII precursor.
DE F8.
OS Fugu rubripes (Japanese pufferfish) (Takifugu rubripes).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Acanthopterygii; Percormorpha; Tetraodontiformes;
OC Tetraodontidae; Tetraodontidae; Takifugu.
OX NCBI_TaxID=31033;
RN [1]
RP SEQUENCE FROM N.A.
RA Davidson C.J., Hirt R.P., Ial K., Snell P., Elgar G.,
RA Tuddenham E.G.D., McVey J.H.;
RT "Comparative sequence analysis and molecular evolution of blood
coagulation genes from Gallus gallus and Fugu rubripes.";
RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
EMBL; AF465279; AAO33374.1; --
DR GO:0005507; F:copper ion binding; IEA.
DR GO:0007155; P:cell adhesion; IEA.
DR InterPro: IPR001117; Cu-oxidase.
DR InterPro: IPR008972; Cupredoxin.
DR InterPro: IPR000421; FA58 C.
DR InterPro: IPR008979; Gal_bind_like.
DR Pfam: PF00394; Cu-oxidase; 2.
DR Pfam: PF00754; FS_F8_type_C; 2.
DR SMART: SM00231; FA58C; 2.
DR PROSITE: PS01285; FA58C_1; 2.

not
print

Ds	1296	LVECTVGEQLAGRAKLVNPNQCSREPLGKSGRIGDSQIKASDVIGNWLPRLARLDS	1355
Qy	2076	GSINAWSTKEPESWIKVDLAPMIHIGITQARQKFSLSYISQFTIMTSLSKQKWQTYR	2135
Ds	1356	GYINAWMGYNESWIKVDLAPMIHIGITQARQKFSLSYISQFTIMTSLSKQKWQTYR	1415
Qy	2136	G-----NGTGLMVFNGVDSGIGKHNLFNPNPIIARYIRLHPHTHSIRSLRMELMGCD	2189
Ds	1416	GSQSSSSSSSTAKVGNLDSRVKNPNFPFVARYIRIHPLYNQRPALRMELMGCD	1475
Qy	2190	LNSCSMPLMGSKAISDAQITASSYTNMFPATWSPKARLHLCGRNANRPQVNNPEWL	2249
Ds	1476	LNSCSPLGLQDRRIPEDSFVASSYSWLLRSWTPSLARLHQSANAWRPQNNPEWL	1535
Qy	2250	QVDFQKTYKVTGVTGQVKSILTMVVKSEFLSSQDGHQWTLFPQNG--KVYVFGNQD	2307
Ds	1536	QVDLGKVKVITGVGTQARSLLTKFMVTFBSVTISRDQGSWVLEGSQREXIFQGNND	1595
Qy	2308	SFTPVVNSLDPPLLTRYLRHPQSWHQIALRMEVLGCEAQ	2348
Ds	1596	SDBEALTIIDAPLFGYIRIHPGLWINDIALRLEVLCGCTQ	1636
RESULT 5			
Q90X47			
ID	Q90X47	PRELIMINARY;	PRT; 2119 AA.
AC	Q90X47		
DT	01-DEC-2001	(TEMBLrel. 19, Created)	
DT	01-DEC-2001	(TEMBLrel. 19, Last sequence update)	
DT	01-DEC-2003	(TEMBLrel. 25, Last annotation update)	
DE	SC:B22015.3	(Novel protein similar to vertebrate coagulation factor V and VII).	
DE	SC:B22015.3		
OS	Brachydanio rerio (Zebrafish)	(Danio rerio).	
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;		
OC	Cyprinidae; Danio.		
OX	NCBI_TaxID=7955;		
RN	[1]	SEQUENCE FROM N.A.	
RP	Lloyd D.;		
RL	Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.		
RC	-/- SIMILARITY: CONTAINS 2 FS/8 TYPE C DOMAINS.		
DR	EMBL; AL590146; CAC94896.1; -		
DR	GO; GO:000507; F:copper ion binding; IEA.		
DR	GO; GO:0007155; P:cell adhesion; IEA.		
DR	InterPro; IPR001117; Cu-oxidase.		
DR	InterPro; IPR008972; Cupredoxin.		
DR	InterPro; IPR000421; FA58 C.		
DR	InterPro; IPR008979; Gal Bind like.		
DR	Pfam; PF00394; Cu-oxidase; 2_		
DR	Pfam; PF00754; F5_F8 type C; 2.		
DR	SMART; SM00231; FA58C; 2.		
DR	PROSITE; PS01285; FA58C_1; 2.		
DR	PROSITE; PS01286; FA58C_2; 1.		
DR	PROSITE; PS50022; FA58C_3; 2.		
DR	PROSITE; PS00079; MULTICOPPER OXIDASE1; 3.		
QY	SEQUENCE	2119 AA; 240643 MW; DCOE806FFA8761E6	CRC64;
Query Match			
Best Local Similarity 30.3%; Pred. No. 2e-159;			
Matches 752; Conservative 379; Mismatches 825; Indels 529; Gaps 68;			
QY	11	LCILRR-CFSAT--RYYLGAVELSDVYMQSDLGELPVDARPPRPVKFPFNTSVYK	66
Ds	13	LALLAFICHCAVERHYIAANINMDVTSQGO-----RTGQSYK	53
QY	67	KTLFVEPTDLHFIARPPPMWGLGPTIOAEVYDTVITLKNMASHPVSLHAGVSYWK	126
Ds	54	KVYREYNEG-FKQKAPHLSSGLLPTLRGQCDTIIVTFRNWADHPCSLHPHGAYGK	112
QY	127	ASGAEYDDQTSOREKDDKVFPGGSHYVQVLEKNGPMASDPLCLTYSYLSHVDLVKD	156

Ds	113	QSEGSYFDNTSLLKNDVDVIQGEHTYQMDVTSQVTTAADPPCITVYLSHFDIVD	172
QY	187	INSGIIGALLVREBSLAKBKQTTLH---KFILLFAVDEGKSWHSETKNSLMQDRDAAS	243
Ds	173	YNTGLIGPLMLIKGGLTDDSGNQ-IHFQESVLLFGVFDENKSWYS-----TGDS	221
QY	244	ARAPQCHTWNVYANRSLPGLIGCHRKSVYHVGMTTPEVESIFLEGHTFLVRNHQA	303
Ds	222	POPLNVKITINYCTNGSVFDDLDICAHSKVSWHLLGMSSEPELFSVHFNQVILLHCHKTS	281
QY	304	SLEISPIITFLTAQTLIMDLGQFLSCHISSHSHQDGMAYVKVDSQPE--EPOLRMKNNEE	361
Ds	282	AVGIISGTATASMTGVHPRWLNGSHISKLEAGLHGYLNIRKDCDYATAPKRLTIEQ-	340
QY	362	ABEDYDDDLTSDMDVVRFDNNSPSFIQIRSVAKKPKTWVHYIAAEEDNDYAPLVLAP	421
Ds	341	-----KKESQEWTYMAAEVIMDYAPNMPEN	367
QY	422	DDRSYKSOYLNNGPQIRGKYKVRFMAYTDTFTFR---EAIQHSGLIGPLLYGEGD	478
Ds	368	MDGDFRSKYLKQGPQIRGKYKAVFTQYKGMFKERADKORKRELGLIGPVIRAIQID	427
QY	479	TLIIIPKQASRPYNIYPHGIITDVRPLYSRRLPKGVKHLKDPFLLPGEIFKYKWTVED	538
Ds	428	IIKIVFKKASRPYSIYPHGLTIIDKAAEGASYPQGN--QTVSQVGETVYTVWSVEED	485
QY	539	GPTKSDPRCLTRYSSFNKMERGLASGLICPLLCYKESVDQDQGNQIMDKENVILFSVP	598
Ds	486	VPTSDPRCLTMTMSAVDAPRDIASGLVGPLLICKSOLNKKVQLKADKQOHAMFTVF	545
QY	599	DENRSWLTENIQRLPNPAGVQLEDPEPQANIMHSINGVYVDSIQ-LSVCLHEVAYWY	657
Ds	546	DENKSYQDENINTVCSDPKVKKDDPEFYKNVWHTINGVYVESQELGFCHEIVTWH	605
QY	658	ILSIGAQDTFLSVFSGYTFKHKMYEDTLTFPESGETVFMMSMENPGLWILGCHNSDFR	717
Ds	606	VSSVGQDYQIQTATFYGHTFELKNEEDILSLFMTGETITKMMVNIIGIWLASLNSHDS	665
QY	718	NRGMTALLKVSQCDKNTGTY-YEDSVEDISAVLLSKNAIEPRSFQNSRHSRSTROKFN	776
Ds	666	TKGMVKFKDLECFR---DYVEYDIED-----GKFTAWKPTI-----	701
QY	777	ATTIPENDIEKTDPMFAHTPMFKIQNVSSDILLMLLROSPTPHGLSLDLQSAKYETFS	836
Ds	702	-----NEIKKEEPEVAR-----	713
QY	837	DDPSGAIDSNLSLSEMTHTFRPLQHHSGDMVFTPESGLOLRINEKIGTTAATLKLKDFK	896
Ds	714	---PDVDEYSDLFAET-----LNLRTFNKVKDEVEIDLTL	746
QY	897	VSTSNLITIPSDNLAAGTNTSSLGPPSPVHVDSQLDITLFGKKSPLTESGGPLS	956
Ds	747	FLDQDDGLLPIVEEKSLGSGNEN-----LINATLQSFIEHNG-----	783
QY	957	LSENNDSKLLSEGLMNSQESSGKNVSVSTESGRLPKRAHGFPALLTKDNALFKVISL	1016
Ds	784	-----LLMEEGDLDKGESSKNVLNDSTK-----	809
QY	1017	LKTNKTSNNS---ATNRKTH---IDGPSL--LIENSPSVWQNLLESDTEFKKVTPLIHOR	1068
Ds	810	LETTTTFDSNRVVALNNETDSIILDFPIVERKVSAPSKPMNEPESTVNTFKTEHI---	866
QY	1069	MLMDKATALLNHNKNTTSKKNEMVQKKGPIPP-DAQNPDMSP-----FKMLFLP	1122
Ds	867	-----NSSLERINAIYSPITETNINTMTHTDPSITPFGSGTGEMTNFLEDDTALLSS	921
QY	1123	ESARWITQRTGKNLSNGQSPFKQLVSLGPKXSEVGQNF-----LSEKNKYVWVK	1173
Ds	922	ESEPLQSNQSENRIAFQBELNAKOGTDVDNNNSVKNQIFKYNVPBGDTLSNKKIQV--	979
QY	1174	GETFKDVGLEKMWFPSSRNLFITNLNHNTHNOEKIKOBEIEKKEILIQENVLPQI	1233

Db 980 -----BEDFVLLSSVFSFSEMT-----TMEYDVSQDITVKGESKETAQS 1018
Qy 1234 HTVCTCKFMKLNFLSTQWVEGSDYCAVAPVLQD-----FRSLK-----DS 1276
Db 1019 QELSSTKTYGEIILSLPDIITAFNLSSSVLRNNSLESNESSNETLFPSSNATFSDS 1078
Qy 1277 TNRTKKHT-----AHFKKG--EENLEGLGNQTIQIIVEKYACTTRISPTNSQCNFVTRS 1330
Db 1079 TNATSSPSSTATFADFNTTFSNATFSDFSNRIQMSDSSNATILSDSNAT-----LSDSS 1134
Qy 1331 KRALKQKRLP--EETELEKRIIIVDTSTQSKNMKHLTPS-----TLTQIDYN 1377
Db 1135 NATLSD-----SSNATLSSSNATLSSNATFSDSNATFSDSNATLFGVSYS 1184
Qy 1378 -----EKEKGATQSPSLDCLTRSHSIPQANRSPFLIAKVSPFSIRPIYLTRVL 1427
Db 1-85 SNTTLDSPLESEMTYILSSANDTI-KSHSEVVSNTSQL-----SSSESTENISL--L 1235
Qy 1428 FODNSSHLPAASYRKKSQGVSESHFLQCAKKNLSAILTLEMTGQREVSGISGTATN 1487
Db 1236 YGSLN-----ASSMKNDSESEBEEVYILNKNH-SEAILTSHL--DQKE-EHWGYESKH 1286
Qy 1488 SVTVKQENTV-----LKPDLKPTSGKV--ELLPKVHI-----YCKDLFTET 1529
Db 1287 ELVHKELPDENKVKVKSANSNKPKLEKKVYQVFKKYGKMKTKSKDYKQPR 1346
Qy 1530 SNGSPGHLDLVEGSLLOCTGCAIKWNPANRPGKVPFLXVATES-AKTPSKLLDLAWDN 1588
Db 1347 SPSFP-----RGFGSVLTGRSRR-----VSEDELTEKPIVIGVPRDFN 1388
Qy 1589 HYGQIKPEEKWSQEKSPKTAFFKKKDTILSLNACENHAIANEGNKPEIBVTWAKQ 1648
Db 1389 DYELIYIPKQDEAD-----PDGL-----DHPE----- 1411
Qy 1649 GRTERLCSQPPVLKRQREITRTTLOSQDEIDYDTISVEMKEDPDYD--EDNQSP 1707
Db 1412 -----EYVYKDPYS---KTADVQALDATSQDILLK 1440
Qy 1708 RSFOKTRHYPIAAVERLWDYMGSSPHVLRNRAQSGSVPOFKKVPQEFDTGDSFTOPLY 1767
Db 1441 MAGDKNTRTYFISVEEBENDYA-GYGORLIDKTAQNERPTVFRKVFERYLDSFISIRDI 1499
Qy 1768 RGEUNELHGLIGPIRAREVDNIWTFNQASRYSYFVSSILSY-----EDQOQA 1819
Db 1500 RGEDEHLGLIGLPIKAEVDQTVVFPFRNARSYSUHGAVKYLKQWGLSYDDESFPYW 1559
Qy 1820 BPRXKFKVNETKTYFWKVQHMAFTKDEPCKAWAYPSVDLAKVHSGLIGLLVCHT 1879
Db 1560 YKQDDAVP2P2NGTFYMTYINPKSGPQNNESSDRTWYYSVAVWPERDINSGLIGLLVCRK 1619
Qy 1880 NTLN--FAHGRQVTVQEFALFTTFDETKSWYFTENMERNCRAPCNQMEDPTPKENYRF 1937
Db 1620 GTLDPKPEDRR-----BFVLFMTFDEKSKWLYENRQRIERKRRVVM-DPNFQDLNKF 1673
Qy 1938 HAINGYIMDTLPLGLWAAQDQIRWYLLSMGSMENIHSHFSGHFTVTKKEEYKVALYNL 1987
Db 1674 DALNG-IYYSKGLRXYTNQAKHNLNMGSPKLNHSHFHGQTFINKELKDRHQGYIPL 1732
Qy 1998 YPGVFTEVMLPSXAGIWRVECLIGELHLAGMSTLF-----LVYSNK 2039
Db 1733 LPPGFATLEMLPSXEGWLQLESEVGLSQOQGMQTLFLFDESYYKYRVQCVLRSYLIYFV 1792
Qy 2040 QOTPLGASGHIRPQFOTASQYQWAPKLARLHYSQSNWST-KPPFSWKVDLLAPM 2098
Db 1793 CDHPLGISGTVQDEQITASRQGWPHLARLHNTGKYNWSTSPBQVQLQVDFQFV 1852
Qy 2099 IIEGKIQGAPKFSLSVISOFILMYSLDGKKWOTYVGNSTG-----LWTFP----- 2146
Db 1853 VISKATQGAQFTHNFVNLITYSYDCKKWIYYKGDSDAVKRNHKTNWFNFARQT 1912
Qy 2147 --GNVDSGKHNIFNPPFIARYIRLPHTHYSIRSLRVELMGCGLNSCMPLGMSKAI 2204
Db 1913 FEGNABAYETKENIFFPPLIGRYVRLHPLSHSYNFPVRLVLEYGCELDGCGCVPLGMEGLI 1972

Qy 2205 SDAQITASSYFTNMFA-TWSPSKARLHLOGESNAWRPQVANNPKWLOVDFQKTKWKTGVT 2263
Db 1973 DSKITASSYASWYSQWHPWRLNKQQTANAWQAKNDIQFWIQVEIKVKKILGIV 2032
Qy 2264 TOGVKSLLTSMYKVEFISSQDGHQWTLFFQNG--KVKVFOGNQDSFTPVVANSLDPPLL 2321
Db 2033 TOGAKNGENMFVRSYILEYSEDGRWMKYTDDEDEQKLFQGNNTDNGIKKVIYPPIF 2092
Qy 2322 TRYLRHPQSWHVCALRMEVLGCE 2346
Db 2093 SRFIRIIPKQWQSVTMRIBELGCD 2117

RESULT 6

O88783 PRELIMINARY; PRT; 2183 AA.
AC O88783
DT 01-NOV-1998 (T-EMBLrel. 08, Created)
DT 01-NOV-1998 (T-EMBLrel. 08, Last sequence update)
DT 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)
DE Murine coagulation factor V.
GN F5.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98282202; Pubmed=9616155;
RA Yang T.L., Cui J., Rehmtulla A., Yang A., Moussalli M., Kaufman R.J.,
RA Ginsburg J.
RT "The structure and function of murine factor V and its inactivation by
RT protein C".
RL Blood 91:4593-4599(1998).
CC -!- SIMILARITY: CONTAINS 2 F5/8 TYPE C DOMAINS.
DR EMBL; U52925; AAC99553.1; -.
DR PIR; T42764; T42764.
DR HSP; P12259; ICZT.
DR MGD; MGI:88382; F5.
DR GO; GO:0005507; F: copper ion binding; IEA.
DR GO; GO:0007155; P: cell adhesion; IEA.
DR InterPro; IPR008117; Cu-oxidase.
DR InterPro; IPR008972; Cupredoxin.
DR InterPro; IPR00421; FA58 C.
DR InterPro; IPR008979; Gal bind like.
DR Pfam; PF00394; Cu-oxidase; 3_type_C; 2.
DR Pfam; PF00754; F5_P8_type_C; 2.
DR SMART; SM00231; FA58C; 2.
DR PROSITE; PS01285; FA58C_1; 2.
DR PROSITE; PS01286; FA58C_2; 2.
DR PROSITE; PS50022; FA58C_3; 2.
DR PROSITE; PS00079; MULTICOPPER OXIDASE1; 2.
SQ SEQUENCE 2183 AA; 247228 MW; BFCBA8AA723F60317 CRC64;

Query Match 22.3%; Score 2774.5; DB 11; Length 2183;
Best local Similarity 29.9%; Pred. No. 2.4e-156;
Matches 750; Conservative 408; Mismatches 851; Indels 503; Gaps 73;

Qy 8 CFELCLL--RFC-----FSATRYLGAVELSDYMQSDLGHLPLVDARPPRPVPS 56
Db 7 CFELVVLGTRWAGWSHQAEAAQLKQFYVAAQGLWNVHE-----PTD 51
Qy 57 FPENTSVYKTLFVFTDHLFNIAKPRPFWMGLLGTPTQAEVYDVTWITLKNMASHVPS 116
Db 52 PLSNLSIPSPFKIIVREY-EQYFKKEPRSNSSLGLTLYAEVGDVYKHFHFNKADKPLS 110
Qy 117 LHAVGSYKASGAEYDDQTSOREKDDKVPFGSGHYVWQVLKNGPMASDPLCLTYS 176
Db 111 IHPQIKYKFSFGASYADHTFPAERKDAVAPGEETVYEWIVSEDSGTPDPPCLTHI 170
Qy 177 YLSEVDLVKDNLSGLICALLVCEGSLAKEKITL--HKFILLFAVEDEGSKSHSETKNS 234


```

QY 2242 VNNPKRMLOVDFQKMKVGTNTQGVKSLTSMYVKEFLISSQQDGHQWTLFFQNGKV-- 2299
D5 2075 ANNKQMLQVLLKIKKTAIVTQCKSLSSGMSYKYSIQVSDQGVAKKPYRQKSSMVD 2134
QY 2300 KVFQCNQDSFPPVNSLDEPLTLRLTRHPSQWQHQIMRNEVLGCEAQDLY 2351
D5 2135 KIFEGSNKCKHMKNFPPPIISIRIIPKTNQSIARLRLFGC--DIY 2183

RESULT 7
Q804W5 PRELIMINARY; PRT; 1802 AA.
AC Q804W5.
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Coagulation factor V precursor (fragment).
GN F5.
OS Fugu rubripes (Japanese pufferfish) (Takifugu rubripes).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
OC Tetraodontidae; Tetraodontidae; Takifugu.
OX NCBI_TaxID=31033;
RN [1]
RP SEQUENCE FROM N.A.
RA Davidson C.J., Hirt R.P., Lal K., Snell P., Elgar G.,
RA Tuddenham E.G.D., McVey J.H.;
RT "Comparative sequence analysis and molecular evolution of blood
RT coagulation genes from Gallus gallus and Fugu rubripes."
RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF465280; AAC33375.1; -.
DR GO; GO:000507; P:copper ion binding; IEA.
DR GO; GO:0007155; P:cell adhesion; IEA.
DR InterPro; IPR001117; Cu-oxidase.
DR InterPro; IPR008972; Cupredoxin.
DR InterPro; IPR000421; FA58 C.
DR InterPro; IPR008979; Gal_Bind_like.
DR Pfam; PF00394; Cu-oxidase; 1.
DR SMART; SM00231; FA58C; 2.
DR PROSITE; PS01285; FA58C_1; 2.
DR PROSITE; PS01286; FA58C_2; 1.
DR PROSITE; PS00222; FA58C_3; 2.
DR PROSITE; PS00079; MULTICOPPER_OXIDASE1; 2.
FT NON TER
FT 1
SQ SEQUENCE 1802 AA; 205160 MW; E2EC3D1D9AA836FD CFC64;

Query Match 21.6%; Score 2688; DB 13; Length 1802;
Best Local Similarity 30.4%; Pred. No. 2.6e-151;
Matches 710; Conservative 322; Mismatches 719; Indels 586; Gaps 67;

QY 62 SVVYKTLFVFTDHLFNIAKPRPPFMGLGFTIQAEVYDVTWITLKNMASHPVSLHVG 121
D5 2 SPTYKVVVREY-DKDFQPKSRFPWLGLLPTLRAEEGEIIVTFRLATKPSIHPHG 60
QY 122 VSYWKAEGARYDQTSQREKEDKVPFGSHYVWVQVLENGPMASDPLCLTYSLSHV 181
D5 61 VAYGQSEGANFYDNTSQEKEDDVWVPNSHRYWBITSDVSPQNDPTCLTYTISHK 120
QY 182 DLVQDLNSGLIGALLVCREGSLAKEKQ--TLRHEILLPAVDFEGKSWHSTKNSLQDR 239
D5 121 DVEEYNSGLIGALLVCKSGSLDESGQIGIYHSEYVLFVGFVENE-----SKPKQN 172
QY 240 DAASARAWPKHTVYNNRSLPGLIGCHRSVYVHWIGMTTPEVHSIFLEGHTFLVRN 299
D5 173 DPASVD--HIKYTINGTEGSLPDVSICTYAPVSLHLVGMSSDEVSFVINGOVLOQNG 230
QY 300 HRQASLEISPTFTATLMDLQCFLLSCHISSHQHDGEAYKVDSCP--EPPQLRMK 357
D5 231 HKMSSVGLSGSSTVSMVAVHTGRWLLSSQIMKHIQGMVGFVNMBCIGCFKAPQRTLT 290
QY 358 NNEAEADYDDDLTDEMDVVRFDNDNSPFIQISVAKKPKTWHVIAAEEEDWDIAPL 417

```

```

D5 291 REQ-----RKSREWTYIAAEIIVANTAPN 316
QY 418 VLAPDDSYKSOYLNNSQPQIRGRYKVKVRPMAYTDTFTK---REAIQHSGLIGPLLYG 474
D5 317 EPAHIDQYKLYLRQSSSTRIGGKKAIVTLTYNESFTQVHKKQKDELGILGPVIRA 376
QY 475 EVGDTLLIIFKNOASRPNIYPHGITDVRPLYSRRRLPGVKHLKDFPLIGEIKPKYKTV 534
D5 377 QIRDVIVIKVKAIRPYSIYPHGLTIKSEGVNYPGSH--SHSVQPGETHYTWKV 434
QY 535 TVEGPTKSDPRCLTRYYSFVNMERDLASGLIGPLLIICYKESVDQGNQIMSKRNVL 594
D5 435 VEDEPLDDDDARCLTRYLSAVIDPFDIASGLIGPLLIICKESGLNVNRVQLRAKQEHAM 494
QY 595 FSVFEDENRNVLTENIORFPLNPAGVQLEDPEFOASNMHSINGYVDSLO--LSVCLHEV 653
D5 495 FAVEDENKSWYLDNI--RYRCEHSKVNKADPDFYKSNWMSINGYVFPSSVPLGFCNGEV 553
QY 654 AYWYILSIGAQDTFLSVFFPGYTPKHKMWYEDTTLTLPFSGETVFMGMENPGLWILGHN 713
D5 554 ATWHVSSVGAQDTYQTATFYGHTFELNERTEDFLSLYPMTGTEISSMMNDNIGWLLASLN 613
QY 714 SDFNRGWTALLKVSSCDKNTGDIYEDSYEDISAYLLSKNNAIEPRFSQNSRHPSTQK 773
D5 614 SHGTTKGRVKFQDVECYD---YQVEYDD-----STNV 644
QY 774 QFNATTIPENDIEKTDWFPAHRTMPKXIQNVSSDDLMLLRQSPFTHGLSISLQBAKYE 833
D5 645 EFNV-----W-----NPLSLDTIQE--- 659
QY 834 TFSDDPSFGAIDSNNSLSEMTFRPQLHSGDMVTFPESGLQLRLNEKLT--TAATLKK 892
D5 660 ---NEKPIVENVK--GEFDSYTEDLAN-----ELGRLSKQOSGSDVQLDSL 705
QY 893 LDFKVSSTSN---NLISTIPSNLAAGTDNTSLGPPSPMPHYDSOLDTTLFGKKSSPL 948
D5 706 LDYEVVDVSNETQINFTKT-KDNDFAAEINGLN-----DT----- 739
QY 949 TSGGPLSLSENNDSKLLESGLMNSBESSGKXNVSTESGRLFKGR-----HGP 1000
D5 740 -----SISNRLDVQNLTGADLLN-QQINWNTVQNNSTLAPLAHNSVNEIKLMHNSP 791
QY 1001 ALLTKONALFKVSIILKTNKTSNATNKRKTHIDGSLIENSPPSVQWVILSDSTEF-K 1059
D5 792 Q--TKNT--FSVIDTDLAAATVNTSA-----LSVDNYSK 824
QY 1060 KVTFLIHDMKMLKATLRLNHNMSKNTTSK--NMEMVQOKKEGPIPPDQNPDMSPFKM 1118
D5 825 EITNLA--GTLOGKNLTA---SDSNVTSVSRINM-----PATRIP----- 859
QY 1119 LFLPESARWQRTQ--GKNSLNSGOGSPKOLVSLGPEKSVGEGNFLSEKKNVWVGKEFT 1177
D5 860 -----SCLSVSKAHVGN--NSSDGTNSSTLEAHGP-----YLNSSGV----- 896
QY 1178 KQVGLKEMVPPSRNLFNTLNLHNNTHNOBKIKQEEIEKKEKTLTIQENWVLPQIHVT 1237
D5 897 -----NPFISKSGNVAALLKNGSVTVKLPMKSEEL----- 927
QY 1238 GTKNFMKNLELLSTRQNVESYDCAVAPVLDPRSLNDSTNRKTKHTAHFKKKEBENLE 1297
D5 928 -----DNTSLKNPSNKTSAESS----- 944
QY 1298 GLGNQTKQIVEXYACTTRISPNTSQCNFVTQSRKALKQFRULPBLETELEKRIIVDDTST 1357
D5 945 -----TPSVSTNDIMTSSKE-----LSSSESSEKVEF----- 971
QY 1358 QWSKNMKHLTPSLTQIDYNEKEKGAIQSPISDCLTRGSHSIPOANRSLPLAKVSSFFPS 1417
D5 972 -----IYVKDKKAGLI----- 982
QY 1418 IRFIYLTRVLFDQNSHPLPAASYRKQSGVQVESH--FLQAKKNNLSAILTLENTGQ 1475

```

Db 983 -----KTSVKTSGHNNWYDGTGK-----IVSAEIPDDM 1011
Qy 1476 REVSGLSGTSATNSVYKKQVNTVLPKDLDP-----KTSKVELLPKVHIYQKDLPTET 1529
Db 1012 KKFEM-SPQNKKKTRVN-----RHRQKGHMKTKRKYKPO-----PRSGPLSP 1061
Qy 1530 SNGSPGHLDLVEGSLLOCTGCAIKWNEANRP-GKVPFLRVATESSAKTPSKLDPLAWDN 1588
Db 1062 RGNP-----LMSPGA-----RPGQLQPVNN--DETLMNPVVIGVPRDPFS 1102
Qy 1589 HYGTOIPKEZKSOEKSPEKTAFFKKDTILSLACSNHAIINAEQNPETIEVTWAKQ 1648
Db 1103 DYELPLGDE----- 1112
Qy 1649 GRTERLCSQPPVLKRHQREITRTTLOSODEIDYDDTISVEMKKEDFDIYDBENQSPR 1708
Db 1113 -----PDHLDVQNVK-----ANEYEVNYKDPYSNEDAKNLHJHQRKYLENR 1158
Qy 1709 SFQKKTTHYFTAAVERLDWYGMSSPHVLNRAGSGS---VPQPKVVVFQFTDGSFTQ 1765
Db 1159 --DKDVRTYFAAEVQWDY-----AGYQRRRRRPPGQNRHTKPKIVVFLYLDSSRTP 1212
Qy 1766 LYRCELNEHLGLIGPYTAAVEUNIMYTRNOASRPYSFYSLISYEBDQOGA--BPRX 1823
Db 1213 EVRGEVDHGLIGLPVKAIEVGQTIMVVFKNASRPFSLHPNGVSYSK-QTEGLSYEDGS 1271
Qy 1824 NF-----VKPNETKTYFWKVQHMAPTKDEPCKAWAYFSDVLEKDVHSGLIGLLV 1876
Db 1272 NYWKYDNEVQGNATFYIWNVPMVGPTDSENCRTWAYISGVNPERDIHSGLIGLLV 1331
Qy 1877 CHNTUNLPAHGRQVT--VOEPALFTTIFDETKSWYFTENMEENCRAPCNIQMEDPTFKENY 1935
Db 1332 CQKGTLN---QELTNTREEMLLFMTDESQSWYFDRNREIMLRNWX-KVMDPDIWENL 1386
Qy 1936 RFHAINCYIMDTPLCLWAOORIRWLLSWGSENTHSFSGHVTIVTKK-BEYKMAL 1994
Db 1387 KPHSINGIITN-LKGLRMYTNQLVSMHLINMGSPKDNVHFHGTTLHKKKTTISYQAV 1445
Qy 1995 YNLYPGVFETVEMLPKAGIRWRECLIGEHLHAGMSTLFLVYGNKCTQPLGMASGHIRDF 2054
Db 1446 YPLLPGSPATLVMYPSKEGLWQLETEVGINECKYQTLFLVLADDCVHPGLRESGVND 1505
Qy 2055 QITASGOYGQWAPKLARLHSGSINASTKEPFWIKVDDLAPMIHIGITQGARQKPS 2114
Db 1506 QITAINTRGYWEPHLARLHNGKTNASTQCNYSWIQVDFQRPVVISQVATQAKQFFQA 1565
Qy 2115 LYISOFITWYSLDGKKWQTYRGNSTGLMVFFGNVDSSGKHLNENPILARYLRLPHT 2174
Db 1566 QYSSYVYISYNDNRWSFYGDSRDDIKVFTGNNDYDVKQNTFFPPLIGRFRIRFHLK 1625
Qy 2175 YSIRSLRMLMGCDLNSCSYPLGMSKASDAQITASSYFTNFA-TWSPSKARLHLQ 2233
Db 1626 WYNKATLRMPFYGCELDGCSVELGMSGLIEDHQITASSASKWYSGTWPFPLGRLNKEG 1685
Qy 2234 RSNARFPQVNPKEWLOVDQTKMKTGVTTOGVKSLLTSMYKVEFLISSQCHQWTLF 2293
Db 1686 TINAWQAKYNDMPQWLQVETQVKITGIVTOGAKFLGAEMFVTSFSLQYSHDGINHPY 1745
Qy 2294 FQNGKV--KVFQGNODSFTPVNSLDPLRLYLRHPOQSVWVHJALRMEVLCEAQ 2348
Db 1746 TDDGCVPAKIPMGNTNNNDVKNWYPIFSRFRIVPKWIGSIPMRNEFLGDCDE 1802

RESULT 8

Q804X3
ID Q804X3 PRELIMINARY; PRT: 1377 AA.
AC Q804X3
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Coagulation factor VIII precursor (Fragment).
GN F8.
OS Gallus gallus (Chicken).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archaeosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RA Tuddenham C.J., Hirt R.P., Lal K., Snell P., Elgar G.,
RA Tuddenham E.G.D., McVey J.H.;
RT "Comparative sequence analysis and molecular evolution of blood
RT coagulation genes from Gallus gallus and Fugu rubripes."
RL Submitted (Jan-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF465272; AAC33367.1;
DR GO; GO:0005507; F: Copper ion binding; IEA.
DR GO; GO:0007155; P: cell adhesion; IEA.
DR InterPro; IPR001117; Cu-oxidase.
DR InterPro; IPR008972; Cupredoxin.
DR InterPro; IPR000421; FA58 C.
DR InterPro; IPR000879; Gal_bind_like.
DR Pfam; PF00394; Cu-oxidase; 1.
DR Pfam; PF00754; F5_F8 type_C; 2.
DR SMART; SM00231; FA58C_2.
DR PROSITE; PS01285; FA58C_1; 2.
DR PROSITE; PS01286; FA58C_2; 1.
DR PROSITE; PS00022; FA58C_3; 2.
DR PROSITE; PS00079; MULTICOPPER_OXIDASE1; 1.
FT NON TER 1
SQ SEQUENCE 1377 AA; 152728 MW; 15ACFD7F765665CD CRC64;

Query Match

Best Local Similarity 36.3%; Score 2511.5; DB 13; Length 1377;
Matches 618; Conservative 210; Mismatches 491; Indels 385; Gaps 48;

Qy 706 LWLGCNSDFRNGMTALLKVSSCDKNTGVDYEDSYEDISAYLLKNNAIERSFS--Q 763
Db 1 VWTLCGNPDRGMAKFTVTQCLE-GSLDEEYEDYE---BEDFLQPRGFSRK 55
Qy 764 NSRHPSTQKQFNATIPENDIEK----TDPFAHRTPEPKIONVS-SSDLLMLLRQSP 817
Db 56 NKRRPCVSEQPNVTS-PKNGTRKPCALCTES--SHEALPHNITNPGSSNGTSAPPGSP 112
Qy 818 TPGLSLSLDQEKYE-----TFSDDPSPGALDSNNSLSENTHFRPOLH 861
Db 113 HPDVTSSLPETTYDFVSYESFLADEEELSKTISQDQAGALPSRK-----H 160
Qy 862 HSDMVFTEPSGLRLNEKLGTTAELKCLDFKVSSTNNLSTIPSDNLAAGTDNTS 921
Db 161 ISGEVGTVSRGLVQSKPAPEADAMVERK-----VTNVL 195
Qy 922 SLGPPSMPVHYDSQLDPTTLFGKKSSPLTSSGGLSLEENNDKLLS-GLMNSQSSWG 980
Db 196 EVQEPK-----KAAVVQAGGTLMLETTETQKPMATYDLWSTVFAAG 238
Qy 981 KVSSTESGLFKGKAHGPALLTKNALPKVISILLKNTKSNNSATNKRTHIDGFSLL 1040
Db 239 K-----GFL-----QCTRSSFQDDGP--- 254
Qy 1041 IENSPSVQWNLDESDFEKKVTEL-IHRRMLMDKNATALRLNHSNKTSSKNWVQOK 1099
Db 255 -EHS-----LGLQVTSSEGADVPLNHESEESNTGPTLGSNHSNFTTDS----- 300
Qy 1100 KEGPIPPDAQNPMSFFKMLFLPESARWIQRTGKNSLNSGQSPKQLVSLGPEKSVEG 1159
Db 301 ---PLGPSARTEDIG-----TSQSHSVESNRSSSELDKLNKRPEKVVSGFNGSLGG 350
Qy 1160 QNPLSEKNVVGKGFTVDVGLKENWFFPSSRNLFNTNLDNLNHNTHNTHQEKIQSE--- 1216
Db 351 ENI-----SFSDLQRVQEQILTDMRNSLPANGSMBEAKGTFAHSDNLEPSRYL 398
Qy 1217 -IEKKTLLIQENVLQIEHTVGTGKFMKNLFLSTQFNVEGSDGAYAPVLQDFSLND 1275
Db 399 PTEERDELILEAV-----FQDATATKOLPDTDSLVLPQSNL----- 434
Qy 1276 STNRTKHTAHFSKKEENLEGLGNQTKQIVEKYACT--RISPTSQQNFVQTSKRAIK 1335

[illegible]

Dc 1354 HFRHWHNHAIURIEFLGCDTQOEY 1377

RESULT# 9
Q7SZNO PRELIMINARY; PRT; 1460 AA.

ID Q7SZNO AC Q7SZNO;
DC Q7SZNO;
AT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Pseudarin C precursor.
OS Pseudonaja textilis (Eastern brown snake).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Lepidosauria; Squamata; Scleroglossa; Serpentes; Colubroides;
OC Elapidae; Acanthophiinae; Pseudonaja.
NCBI_TaxID=8673;
RX [1]
RN SEQUENCE FROM N.A.
RP TISSUE=Venom gland.
RC MEDLINE=22781770; PubMed=12730119;
RAO Rao V.S., Swarup S., Kini R.M.;
RT "The nonenzymatic subunit of pseudarin C, a prothrombin activator from eastern brown snake (*Pseudonaja textilis*) venom, shows structural similarity to mammalian coagulation factor V.";
RL Blood 102:1347-1354(2003).
RE EMEL; AY168281; AAC038805.1; -.
KW Signal.
FT SIGNAL.
SQ SEQUENCE 1460 AA; 165931 KW; 6AFB63E2D5D275A6 CRC64;

Query Match 18.8%; Score 2340; DB 13; Length 1460;
Best Local Similarity 25.6%; Pred. No. 1.2e-130;
Matches 601; Conservative 287; Mismatches 511; Indels 952; Gaps 44

Qy 22 RRYVLGAVELSDWK--NQSDLGELFVDPARPPRPVKSPFPNTSVVYKTLFEVFTHLFN 79
Db ||:||:|||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Qy 80 IAKRPPWMLGLGTIOAEVVDTVTILKNASHPVSLHAVGVSWKASGEABYDDOTSQ 139
Db ||:||:|||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 78 QEERDALSGLGPTLRGEVDSLIIFYKFNFATQPSVIHPQSAYNKWSESSVS DGTSD 137
Qy 140 REKEDDKVPFGGSHTYVQVUKENGMASDPLCTYISLVSHVDLVKDNLNSGLIGALVCR 199
Db ||:||:|||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 138 VERLDADVPGQSPKYWNMTAEITGPKKADPPCITYAYISHVNVMVDFNSGLIGALTICK 197
Qy 200 EGSLAKEKTOTL--HKTIILFAVDEGSKHSSETKNSIMQDRDAASARA WPKMHTVNGYV 257
Db |||||:|||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 198 EGSUNANGSKQFFNRVFLMFVSFDESNNWT--RKPSL-----QTINGFA 241
Qy 258 NRSLPGLIGHRKRSYVHVHVGMTTPPVHSHIFLEHTFLVRNHRQAASLEISPIFLTAQT 317
Db ||:||:|||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 242 NGTLPDVQAACAYCHISHLIGMSSPSEIFS VHENGQTLEQHNYKVSTINLVGGASVTADM 301
Qy 318 LLMDLGOFLLSCHSISSHCHDMEAVKVDSCPPEQLRMKNNEAEADYDDDLTDSEMDEV 377
Db ||:||:|||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 302 SVSRTGKLWLSLVAKHLQAGMYGLMKCGNDPTLTRK----- 341
Qy 378 RFDDNDSFSFTQRSAVKKPKTWVHYIAAEEEDWDVAPLVLPDDRYSKSYLNNGFQR 437
Db ||:||:|||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 342 -----LSPRELMKI-----KNWEYFIAAEITWDYAPEIPSPVDSRYKAQYLDNFSN 389
Qy 438 IGKYYKKVRFWAYTDETTF-KTRERAI--OHESGILGPLLYXGEVGDILLIIFKNQASRPNIY 495
Db ||:||:|||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 390 IGKYYKAAVPRQVEDGNFTPTPAIWPKRGKILGPVIKAKVRDPTVIIFFKNLASRPSIY 449
Qy 496 PHGITDVS---PLYSRRLPKGVAKHAKDFPLLCEIFPKYKWTIVTEGPKSDRCPLTRY 551
Db ||:||:|||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 450 VHGVSVSKDAEGALYPSDPKENITHKG--AVEPQQVITYKWTVLDTDEPTVKQSECITKL 507
Qy 552 YSSPWNMERDLASGLIGPLLCYKESVDQRGNQIMS DKNRVILFSVPDENRSWLVTENIQ 611
Db ||:||:|||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 508 YHSADVMTRIIASGLIGPLLVCKFKALSIVGXVONKADVEOHAFVAFDENKSWYFENIK 567

Matches	643;	Conservative	379;	Mismatches	779;	Indels	721;	Gaps	80;
QY	86	PMKGLIGTQAEVYD	VVITLKNMASPVSLHVGVS	---	---	---	---	---	124
Db	46	PSLGLLGLTLYAEVGD	TIKVHFNKADKPLSIHPGIRYSFSEAOQLFAEEMDFAAFFI	105					
QY	125	---	---	---	---	---	---	---	124
Db	106	LLRCLFDPPSLSQALVLLPKRKTFYVTPLL	EGVPGVHRLSMGLYTMKRGITFLGQTG	165					
QY	125	---	---	WK	---	---	---	---	139
Db	166	PWCPRSRLHAGAGCDDRGGBEPERASAPFLHT	GENWKTLDVFVAVTPSGASYLDTSF	225					
QY	140	REKEDKVPFGSGTYVWQVLKENGPMASDPLCT	YSLYSLSHVDLAVKDLNSGLIGALLAVCR	199					
Db	226	AERKDDAAGAEETYEWFIEGDSGFPDDP	PCLIHIYSYENLUTQFNSGLIGLPLICK	285					
QY	200	BGSLAKEKTQIL--HKFILLFAVFDGKSWHSET	KNLSMQRDAASARWPKMHTVNGYV	255					
Db	286	EGTLTEGDTQKVFQKQHVLLFAVFDKSW	---	---	---	---	---	---	329
QY	258	NRSPLPGLIGCHRKSVYHVICMGITPEVHSTFL	GHTFLVNHQASLE	---	---	---	---	---	309
Db	330	NRTMPDVTVCAYDVSHLIGNSGPELFSHFNG	QVLEQKQKSVSTLLQTVGPIRP	369					
QY	310	ITFLTAQTLMLDLQFLILSHISHQHDGMEAYV	KVDSCEPQLRMKNNEAEVDDDL	369					
Db	390	IALNTVQYIASRGEDLLVPAT--RAAGMQAY	IDIKNCPK	---	---	---	---	---	429
QY	370	TDSEMDVVRDDDNSPFTQIRSVAKKHPTXVHY	IAAEREDMDYAPLVLA--PDDRSYKS	428					
Db	430	TRSSKTLTR	---	---	---	---	---	---	475
QY	429	QYLNNGPQIRIGKVKYKVRFWAYDTEKTREA	QHSGLLPILLYGVGDTLIIIFKNQA	488					
Db	476	SPMGWEDDNNRN--NGSRMLATVNP	CSRLRSHQYKLGAL	---	---	---	---	---	518
QY	489	SRPNIYPHGITDVRPLYSRRLPKGVKHLKDF	PLPGEIEFKYKWTVVEDGPKSPRCL	548					
Db	519	---	---	---	---	---	---	---	533
QY	549	TRYSSFWNMERDIASGLLPILLYKESVDQR	GNQIMSKRNVILFSVDENRSWLTE	608					
Db	534	VVRHMPFL	---	---	---	---	---	---	565
QY	609	NIQRFLENPAGVOLEDPEQASNMHSINGYVDS	LQ--LSVCLHEVAYVILSIGAQTF	667					
Db	566	NINKFCENPEVKRDPKFTYESNMNTINGYV	DESITLGFCDPDAVQHFCVSGVTHDDI	625					
QY	668	LSVFFSGYTFKRWYEDTILTFSGEYTFVSMEN	PGLWILGCHNSDFNRNMTALLKV	727					
Db	626	LTVHFTGHSPIYGRRHEDTLTFPMGSGSVTV	YMDNIGITWMLTMSNPFRRNLRF	685					
QY	728	SSCKMTGYVYRVEDISAYLLSKNAIEPRFS	QNSRHPSTROKFNATTIPEND--I	785					
Db	686	VKNRNDDD--DEDSYF--IY	---	---	---	---	---	---	725
QY	786	EKTDPPFAHRTPMEKYNSSDLLMLRQS---	PTPHGLSLSDIEAKYVFSDPSG	842					
Db	726	ENEDDDYQ	---	---	---	---	---	---	777
QY	843	AIDSNNS--LSEMTFRPQLHHSQDVFTPE	SGIQLRNLKLTAAATELKKDKFVAST	900					
Db	778	VWDSKSRNLXIT	---	---	---	---	---	---	804
QY	901	NNMLISTIPSDNLAAGTNTSSLGPPSPVHV	DSQDLDTFLFGKKS	---	---	---	---	---	953
Db	805	SGATIAIGILLGNL--TGLGRNSVLN--PSTE	YHSSSYVENDMDEPQSNITWVILPLGAKG	--861					
QY	954	PLSLSENNDKLLIESGLMNSQES--SWGKN	VSSTSGRLFKGKAHGPALLTKCNALFK	1011					
Db	862	--SGSREQTKPKTIKTRZHRMKHFSNMK	---	---	---	---	---	---	892


```
QY 705 GLWILCHNSDFRNGMTALLKVSSCDKNTGDTYYEDSYEDISAYLLSKNNAIPRPSQN 764
Db 692 GIPEIYCQASHREGMOATYNSQC-----SSHQD-----SPRQHYQA 730
QY 765 SRHPSTRQKQFNATTIPENDIEKTDWFAHRTWPKIQNVSSDDLMLLRQSTFPLGLSL 824
Db 731 SR-----732
QY 825 SDLOEAKYETFDSPGAGIDSNNLSLSEMTFRPOLHSHGDMVFTPEGQLRLNEKIGT 884
Db 733 -----732
QY 885 TAATELKCLDFKVSSISNNLISITPSDNLAAGTNTSSLGPPSPMVFHYQSOLDTTLFGKK 944
Db 733 -----VYV-----735
QY 945 SGPLTBSGGPLSLSEENNDSKLLSGLMNSQESSMGKNVSSSTESGRLFKGKRAHGALLLT 1004
Db 736 -----IMAELEW-----743
QY 1005 KDNALPKVISILKTNKTNNSATNRKTHIDGSPSLIENSPPVWQNILSDTEFKKVTPL 1064
Db 744 -----DYCPDRSWELE-----WNTSEKD-----762
QY 1065 IHDRMLMDKNATLRLNHSNKTSSKNMEMVQOKKEGIPDDAQNPDMSPFFMLPLPES 1124
Db 763 -----762
QY 1125 ARWIORTHGKNSLNSGQPPKQLVSLGPEKSVEGQNTLSEKNVVGKGETKOVGLKE 1184
Db 763 -----762
QY 1185 MYFPSSRNLFJNLNHNHNTHNOEKKTOEIERKETLIOENVVLPQIHTVTGTNPFMK 1244
Db 763 -----SYGHVFLSNKO-----773
QY 1245 NLFPLLSTRNGVEGSDGAYAPVLQDPRSLNDSTNRKTHAHSKKGSENLGLGNQTK 1304
Db 774 -----773
QY 1305 QIVEKYACTRISPTNTSQONFVQTSRKALKQPLPLEETELEKRIIVDDTSTQSKNMK 1364
Db 774 -----773
QY 1365 HLTPSTLTQIDYNEKEKGAITOSPDLSTLRSHSIPOANRSPPIAKVSSPPIRYLT 1424
Db 774 -----773
QY 1425 RVLFDNSSLPLPAASYRKDSGVQESSHFLQAKKNLSLAILILEMTGDQREVCSLGT 1484
Db 774 -----GLLGS-----778
QY 1485 ATNSVTYKKVENTVLPKPLPNTSGKVELLPKVHYQKDLFTETNSGPHLDLVEGSL 1544
Db 779 -----778
QY 1545 LQTEGAIKWNENRPGKVPFLRVATESAKTPSKLLDPLADNHYGTQIPKEWKSQEK 1604
Db 779 -----778
QY 1605 SPEKTAKKKDTILSLNACENHAIAINEGQNKPEIEVTWAKQRTRELCSQPPVLKR 1664
Db 779 -----778
QY 1665 FQREITRITLOSDEIDYDDTISVEMKEDFDIYDEENGSPRSFQKKTRHYPIAAVER 1724
Db 779 -----778
QY 1725 LWDYCMSSSPHLNRAGSGSVOPKVVQEFDTGSETQPLYGELNEHGLJLGPYIRA 1784
Db 779 -----RYKKAVFREYTDGTFRIIPRPSGPEHGLIGLPLIRG 815
QY 1785 EVEDNIMVTPRNQASRP2SFYS-SLISVEEDQROGAERKKNFVKPNETKITFWKQJHMA 1843
```

```
Db 816 EVGDIITVVFKNKASRPYSIHAGVLESNTGGQARZ-----PGEVITYQWNIERSG 868
QY 1844 PKDEPCKAWAYFSDVLEKQVHSLGLOPLVCHTNTLNPAHGRQVTVQEFALFTTFD 1903
Db 869 PGPDSACSACSWIYISAVDPIKDWYSLGVPLVICRNGILEPNGRNDMDREFALLFLFD 928
QY 1904 ETKSWYFTENMER-NCRAPCNTQMEDPTKENRYRPHAINGYIMDTPLGLVMAQDQIRWY 1962
Db 929 ENQSWLTKENIATYGPQESSHVNLKDATFLEGNKHAINGKLYANLRGHTVYQGERVAVY 988
QY 1963 LLSMGSNENIHSIHFSGHVTVRKKEEYKMAVLYNLYPGVFTVEMLPSPKAGIWRVECLIC 2022
Db 989 MLAMGQDTHIVHFAESFLYQNGQSYADVVDLFPGTFFEVVEMVASPGIWLMECHVT 1048
QY 2023 EHLHAGMSTLFLVYSNK 2039
Db 1049 DHVHAGMETIFTVLSE 1065
```

RESULT 14
Q9Z0Z4
ID Q9Z0Z4 PRELIMINARY; PRT; 1157 AA.
AC Q9Z0Z4;
DT 01-MAY-1999 (TREMELrel. 10, Created)
DT 01-MAY-1999 (TREMELrel. 10, Last sequence update)
DT 01-OCT-2003 (TREMELrel. 25, Last annotation update)
DE Hephaestin.
GN HEFH.
OS Mus musculus (Mouse).
OC Zukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]_TaxID=10090;
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J;
RX MEDLINE=99140771; PubMed=9888272;
RA Vulpe C.D., Kuo Y.M., Murphy T.L., Cowley L., Askwith C., Libina N.,
RA Gitschier J., Anderson G.J.;
RT "Hephaestin, a ceruloplasmin homologue implicated in intestinal iron
RT transport, is defective in the sla mouse.";
RL Nat. Genet. 21:195-199(1999).
DR EMBL; AF082567; AAD16035.1; -.
DR HSSP; P00450; 1KCM.
DR MGD; MGI:1332240; Heph.
DR GO; GO:0005507; F:copper ion binding; IEA.
DR InterPro; IPR001117; Cu-oxidase.
DR InterPro; IPR006972; Cupredoxin.
DR InterPro; IPR002355; Cu ox copper_BS.
DR Pfam; PF00394; Cu-oxidase; 3.
DR PROSITE; PS00079; MULTICOPPER_OXIDASE1; 3.
DR PROSITE; PS00080; MULTICOPPER_OXIDASE2; 1.
SQ SEQUENCE 1157 AA; 129692 MW; 4119BIDA2BFEC2FD CRC64;

Query Match 11.2%; Score 1388; DB 11; Length 1157;
Best Local Similarity 20.0%; Pred. No. 6.2e-74;
Matches 412; Conservative 179; Mismatches 414; Indels 1052; Gaps 26;
QY 20 ATRRYYLGAVELSWDNYQSDLGELPVDARPPRPVKSF----PNTSVVYKKTFLVBFTD 75
Db 24 AIRNYLIGIQDMQNYAPKGRNVITNQTNNDTVASSFLKSGKNRIGSSYKKTVKEYSD 83
QY 76 HLFNIAKPRPPMGLGTTQAEVYDVTWITLKNWASHPVSLHAVGYWYKASGAGYDD 135
Db 84 GTYTEETAKPAWLGFLGLLQAEVGVILLHLKNAFASRPYTHPHGVYFVKDSGSGYPD 143
QY 136 QTSOREKDDKVPFGSGSHYVQVLKENGPMASDPLCLITYSYLSHVDLVKDLNGLICAL 195
Db 144 GSSGLKADSDVPFGSGSHVYVNSIPESHAPTEADPACLTWIYHSHVDAPDIATGLIGPL 203
QY 196 LVCHREGSJ----AKEKTQTLHKFTLLFAVFDEGKSWHSE-----TKNSLMQDRDAASAR 245
Db 204 ITCRGTLDGNSPPQRKXDVDFNFFLLFSVIDENLSWHLDDNIATYCSDPASVDKEDGAFQ 263

[illegible]

QY	1957	QIRWYLSMGSNENIHSIHFSCHVFTVRKKEFKVLYNLYPGVZETVEMLPKAGIWR	2016
Db	983	ERVAVMYLWYGGQTDIHTVHFHAESEFLYQNGHSYRADVDLFFGTFFVEMVANSNPGAWL	1042
QY	2017	VECLIGEHFHAGMSTFLFYVSNK	2039
Db	1043	MECHVTDHVEAGMETFTVLSHE	1065

Search completed: April 13, 2004, 14:07:51
Job time : 103.5 secs

GenCore version 5.1.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 13, 2004, 14:01:28 ; Search time 22.5 seconds
(without alignments)
5394.340 Million cell updates/sec

Title: NP000123-328
Perfect score: 12416
Sequence: 1 MQIELSTCFLLRFCFSFA.....WVHQIALRMEVLGCEAQDLY 2351

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents_AA*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pdp:*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pdp:*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pdp:*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pdp:*
5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pdp:*
6: /cgn2_6/ptodata/2/iaa/backfiles.pdp:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Length	ID	Description
1	12410	100.0	2351	4	US-10-133-907-4
2	12404	99.9	2351	6	Sequence 4, Appli Patent No. 5171844
3	12401	99.9	2351	1	Sequence 2, Appli Patent No. 5171844
4	12397	99.8	2351	1	Sequence 2, Appli Patent No. 5422260
5	12388	99.8	2351	6	Sequence 4, Appli Patent No. 5422260
6	12293	99.0	2332	1	Sequence 2, Appli Patent No. 5422260
7	12293	99.0	2332	1	Sequence 4, Appli Patent No. 5422260
8	12293	99.0	2332	1	Sequence 2, Appli Patent No. 5422260
9	12293	99.0	2332	1	Sequence 4, Appli Patent No. 5422260
10	12293	99.0	2332	1	Sequence 2, Appli Patent No. 5422260
11	12293	99.0	2332	3	Sequence 2, Appli Patent No. 5422260
12	12293	99.0	2332	4	Sequence 2, Appli Patent No. 5422260
13	12293	99.0	2332	5	Sequence 2, Appli Patent No. 5422260
14	12293	99.0	2332	5	Sequence 2, Appli Patent No. 5422260
15	12293	99.0	2332	5	Sequence 2, Appli Patent No. 5422260
16	12286	99.0	2332	3	Sequence 2, Appli Patent No. 5422260
17	12243	98.6	2332	3	Sequence 2, Appli Patent No. 5422260
18	9485	76.4	2343	3	Sequence 2, Appli Patent No. 5422260
19	8827	71.1	2319	1	Sequence 2, Appli Patent No. 5422260
20	8827	71.1	2319	1	Sequence 2, Appli Patent No. 5422260
21	8827	71.1	2319	2	Sequence 2, Appli Patent No. 5422260
22	8827	71.1	2319	3	Sequence 2, Appli Patent No. 5422260
23	8827	71.1	2319	4	Sequence 2, Appli Patent No. 5422260
24	8827	71.1	2319	4	Sequence 2, Appli Patent No. 5422260
25	8827	71.1	2319	5	Sequence 2, Appli Patent No. 5422260
26	8722	70.2	2304	3	Sequence 2, Appli Patent No. 5422260
27	8349	67.2	1661	2	Sequence 2, Appli Patent No. 5422260

28	8349	67.2	1661	2	US-08-558-107-2	Sequence 2, Appli
29	8349	67.2	1661	3	US-09-243-539-2	Sequence 2, Appli
30	8068	65.0	2133	2	US-08-670-707A-37	Sequence 37, Appli
31	8068	65.0	2133	3	US-09-037-601-37	Sequence 37, Appli
32	8068	65.0	2133	4	US-09-315-179-37	Sequence 37, Appli
33	8068	64.9	2133	4	US-09-523-656-30	Sequence 30, Appli
34	7951.5	64.0	2115	3	US-09-324-867-5	Sequence 5, Appli
35	7392	59.5	1471	1	US-08-683-8393-3	Sequence 3, Appli
36	7219	58.1	1438	1	US-09-203-916-1	Sequence 1, Appli
37	6216	50.1	1467	4	US-09-523-656-38	Sequence 38, Appli
38	6114	49.2	1443	2	US-08-670-707A-39	Sequence 39, Appli
39	6114	49.2	1443	3	US-09-037-601-39	Sequence 39, Appli
40	6114	49.2	1443	4	US-09-315-179-39	Sequence 39, Appli
41	2774.5	22.3	2183	3	US-08-746-111-5	Sequence 5, Appli
42	2436.5	19.6	541	1	US-08-121-202-4	Sequence 4, Appli
43	2258	18.2	435	3	US-08-448-722A-5	Sequence 5, Appli
44	2258	18.2	435	3	US-08-189-309B-5	Sequence 5, Appli
45	2190.5	17.6	868	1	US-07-864-004B-6	Sequence 6, Appli

ALIGNMENTS

RESULT 1
US-10-133-907-4
; Sequence 4, Application US/10133907
; Patent No. 6677369
; GENERAL INFORMATION:

; APPLICANT: Chien, Kenneth R
; TITLE OF INVENTION: Method to treat hemophilia by hepatic gene transfer of Factor VIII
; TITLE OF INVENTION: Method to treat hemophilia by hepatic gene transfer of Factor VIII
; FILE REFERENCE: 6627-PAL170
; CURRENT APPLICATION NUMBER: US/10/133.907
; PRIOR FILING DATE: 2002-04-25
; PRIOR FILING DATE: 2001-04-25
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 4
; LENGTH: 2351
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-133-907-4

Query Match	100.0%;	Score	12410;	DB	4;	Length	2351;
Best Local Similarity	100.0%;	Pred. No.	0;				
Matches	2350;	Conservative	0;	Mismatches	1;	Indels	0;
						Gaps	0;
Qy	1	MQIELSTCFLLRFCFSFA	TRRYLGA	VELSWD	YMQSD	LGELP	VDARFP
Db	1	MQIELSTCFLLRFCFSFA	TRRYLGA	VELSWD	YMQSD	LGELP	VDARFP
Qy	61	TSVYKKTLLFVEFTDHL	FNIAKPP	PMWGLI	PTQAEV	YDVIT	VITLKN
Db	61	TSVYKKTLLFVEFTDHL	FNIAKPP	PMWGLI	PTQAEV	YDVIT	VITLKN
Qy	121	GVSWKASEGAEYDDQTS	OREKEDD	KVPFG	SHYVQ	WIKENG	PMSDPL
Db	121	GVSWKASEGAEYDDQTS	OREKEDD	KVPFG	SHYVQ	WIKENG	PMSDPL
Qy	181	VDLVKLNGLIGALLVCR	EGSLAKE	KTOTL	HLFILL	FAVDE	GKSHST
Db	181	VDLVKLNGLIGALLVCR	EGSLAKE	KTOTL	HLFILL	FAVDE	GKSHST
Qy	241	AASARAWPKMHTVNGY	VNRSPL	GLIGCH	RSVYV	HWI	MGTTPE
Db	241	AASARAWPKMHTVNGY	VNRSPL	GLIGCH	RSVYV	HWI	MGTTPE
Qy	301	QASLHISPTITFTTAQ	TLMDL	QOFLI	LSCHIS	SHQHD	GMEAY
Db	301	QASLHISPTITFTTAQ	TLMDL	QOFLI	LSCHIS	SHQHD	GMEAY

361 EAEDYDDDLTDSBMDVVRFDKSPSFIQIRSVAKKHPKTWVHYIAAEEDBDWVAPLVL 420
Db
361 EAEDYDDDLTDSBMDVVRFDKSPSFIQIRSVAKKHPKTWVHYIAAEEDBDWVAPLVL 420
QY
421 PDRSYKSOYVNNQPGORIGKVKVAFMAYDTDFTKTREAIOHESGILGPLLXGEVGDTL 480
Db
421 PDRSYKSOYVNNQPGORIGKVKVAFMAYDTDFTKTREAIOHESGILGPLLXGEVGDTL 480
QY
481 LIIFKQASRPNYIYPHGIDTVRPLSRRLPKGVGHLKDPPLPGRIFFKKTWTVEDGP 540
Db
481 LIIFKQASRPNYIYPHGIDTVRPLSRRLPKGVGHLKDPPLPGRIFFKKTWTVEDGP 540
QY
541 TKSDPRLTRYSSFVNMBERDLASGLIGPLLI CYKESVDQGNQIMSDKRNVLFSVDFE 600
Db
541 TKSDPRLTRYSSFVNMBERDLASGLIGPLLI CYKESVDQGNQIMSDKRNVLFSVDFE 600
QY
601 NRSWYLITENIQRELPNAGVQLEDDEPQASNIHVSINGYVFDLSQISVCLHEVAYWYLS 660
Db
601 NRSWYLITENIQRELPNAGVQLEDDEPQASNIHVSINGYVFDLSQISVCLHEVAYWYLS 660
QY
661 IGAQTDFLSVFRSGYTFKHXWVYEDTLTLRPPSGETVEMSNPGLWILGCHNSDPRNG 720
Db
661 IGAQTDFLSVFRSGYTFKHXWVYEDTLTLRPPSGETVEMSNPGLWILGCHNSDPRNG 720
QY
721 MTALLKVSXCDKNTGYEDSYEDI SAYLLSKNAIEPRSFQNSRHPSTROKQFNATTI 780
Db
721 MTALLKVSXCDKNTGYEDSYEDI SAYLLSKNAIEPRSFQNSRHPSTROKQFNATTI 780
QY
781 PENDIEKTDWFAHRTMPKPIQNVSSDILLMLRQSPTHGLSLSD:QBAXYTFPSDDPS 840
Db
781 PENDIEKTDWFAHRTMPKPIQNVSSDILLMLRQSPTHGLSLSD:QBAXYTFPSDDPS 840
QY
841 PGAIDSNNSLSEMTHERPOLHSGDMVFTPEGLQLRLNEKLGTTAATLKKLDEKVSST 900
Db
841 PGAIDSNNSLSEMTHERPOLHSGDMVFTPEGLQLRLNEKLGTTAATLKKLDEKVSST 900
QY
901 SNNLIITIPSDNLAAGTNDTSSIGPSPMPVHYDSQDITTLFGKKSPLTSEGGLSLSE 960
Db
901 SNNLIITIPSDNLAAGTNDTSSIGPSPMPVHYDSQDITTLFGKKSPLTSEGGLSLSE 960
QY
961 NNDKILLESGLMNSOSSWGKNVSSYTESGRLFGKRAHGPAITKONALFKVISILIKTN 1020
Db
961 NNDKILLESGLMNSOSSWGKNVSSYTESGRLFGKRAHGPAITKONALFKVISILIKTN 1020
QY
1021 KTSNNSATNRKTHIDGFSILLIENSPPVQNIIESDTEFKKVTPLIHDRMLXDNATLAL 1080
Db
1021 KTSNNSATNRKTHIDGFSILLIENSPPVQNIIESDTEFKKVTPLIHDRMLXDNATLAL 1080
QY
1081 NMSNKTTSKKNMVMQOKKEGIPDPAQNPDMSPFKMLFLPSARWIOETHGKNSLNG 1140
Db
1081 NMSNKTTSKKNMVMQOKKEGIPDPAQNPDMSPFKMLFLPSARWIOETHGKNSLNG 1140
QY
1141 QGPPSPQOLVSLGPEKVEGONFLSEKNVVGKGEFTKOVGLKEMVFPSSRNLFLLNLDN 1200
Db
1141 QGPPSPQOLVSLGPEKVEGONFLSEKNVVGKGEFTKOVGLKEMVFPSSRNLFLLNLDN 1200
QY
1201 LHENNTNHEKKTQEBIEKKEKTELQENNVLPQIHTVTGKNFMKNFLFLSTRONVEGSD 1260
Db
1201 LHENNTNHEKKTQEBIEKKEKTELQENNVLPQIHTVTGKNFMKNFLFLSTRONVEGSD 1260
QY
1261 GAYAPVLQFRSLNDSTNRTKHTAHFSPKSGEENLEGONQTKQIVEKYACTRISPN 1320
Db
1261 GAYAPVLQFRSLNDSTNRTKHTAHFSPKSGEENLEGONQTKQIVEKYACTRISPN 1320
QY
1321 SQONFVTSKRALKQFRPLBETLEKRIIVDDTSTQSKNKHITPSTLTQIDYNEKE 1380
Db
1321 SQONFVTSKRALKQFRPLBETLEKRIIVDDTSTQSKNKHITPSTLTQIDYNEKE 1380
QY
1381 KGAITQPSLSDCLTRSHSIPQANRSLPZAKVSSPFSIRPIYLRVLFQDNSSHLPAASY 1440
Db
1381 KGAITQPSLSDCLTRSHSIPQANRSLPZAKVSSPFSIRPIYLRVLFQDNSSHLPAASY 1440
QY
1441 RKDQSVQSSSHFLQAGAKNNLSLAITLMTGDQREVSLGTSATNSVTYKKVENTVLP 1500

RESULT 2

5171844-2

; APPLICANT: VAN OOVEN, ALBERT J. J.; PANNEKOEK, HANS; VERBEET,
; MARTINUS P.; VAN LEEN, ROBERT W.
; TITLE OF INVENTION: PROTEINS WITH FACTOR VIII ACTIVITY
; PROCESS FOR THEIR PREPARATION USING GENETICALLY-ENGINEERED
; AND PHARMACEUTICAL COMPOSITIONS CONTAINING THEM
; NUMBER OF SEQUENCES: 12

1441 RKDQSVQSSSHFLQAGAKNNLSLAITLMTGDQREVSLGTSATNSVTYKKVENTVLP 1500
QY
1501 KPDLPTKSGVLLPKVHIYQKDLFPFTETSNQSPGHLDLVEGSLLOQTEGAIKWEANRP 1560
Db
1501 KPDLPTKSGVLLPKVHIYQKDLFPFTETSNQSPGHLDLVEGSLLOQTEGAIKWEANRP 1560
QY
1561 GKVPFLRVATATESAKTPSKLLDPLANDNHYGTQIPKEEWKSOEKSPKTAFFKKOTILSL 1620
Db
1561 GKVPFLRVATATESAKTPSKLLDPLANDNHYGTQIPKEEWKSOEKSPKTAFFKKOTILSL 1620
QY
1621 NACESNHAIAANEGONKPEIEVWAKQGTERTLSONPVLKXHQREIRTRITLQSDQBE 1680
Db
1621 NACESNHAIAANEGONKPEIEVWAKQGTERTLSONPVLKXHQREIRTRITLQSDQBE 1680
QY
1681 IYDDTISVEMKKE3PDIYDEBENQSPRSFQKTRHYFAAVERLWDYQMS5SPHYLRNR 1740
Db
1681 IYDDTISVEMKKE3PDIYDEBENQSPRSFQKTRHYFAAVERLWDYQMS5SPHYLRNR 1740
QY
1741 AQSGSVQPKKVVOEFQTDGSEFTQPLRYGELNEHLLGCGYIRAEVEDNIMVTFRNOASR 1800
Db
1741 AQSGSVQPKKVVOEFQTDGSEFTQPLRYGELNEHLLGCGYIRAEVEDNIMVTFRNOASR 1800
QY
1801 PYSFYSSLLSYEBEDQROGAEPKKNFVKPNETKTYFWKVQHMAPTKDEDFCKAWAYFSV 1860
Db
1801 PYSFYSSLLSYEBEDQROGAEPKKNFVKPNETKTYFWKVQHMAPTKDEDFCKAWAYFSV 1860
QY
1861 DLEKOVHSLGILGFLVCHTNTLNPAHGRQVTVQEPALFPTIFDETKSWYFTENMERNCSA 1920
Db
1861 DLEKOVHSLGILGFLVCHTNTLNPAHGRQVTVQEPALFPTIFDETKSWYFTENMERNCSA 1920
QY
1921 PCNIQMEDTFRKENVRFHAINGIIMDTPLGLVMAODQRIWYLLSMG5NENIHS7HFSCH 1980
Db
1921 PCNIQMEDTFRKENVRFHAINGIIMDTPLGLVMAODQRIWYLLSMG5NENIHS7HFSCH 1980
QY
1981 VFTVRKKEBYKMALYNLYFGVFTVEMLP5SKAGIWRVECLIGEHLHAGMSTLFLVYSNKC 2040
Db
1981 VFTVRKKEBYKMALYNLYFGVFTVEMLP5SKAGIWRVECLIGEHLHAGMSTLFLVYSNKC 2040
QY
2041 QTFPLGMA5GHIEDFQITASGQYQWAPKLARLHYG5SINAWSTKBPFSWIKVDLLAPMII 2100
Db
2041 QTFPLGMA5GHIEDFQITASGQYQWAPKLARLHYG5SINAWSTKBPFSWIKVDLLAPMII 2100
QY
2101 HGKIQCARQKPS5SYISOIIMYSLDGKKWOTYRGNSTGTLMVPGNVDSSGIXHNIFN 2160
Db
2101 HGKIQCARQKPS5SYISOIIMYSLDGKKWOTYRGNSTGTLMVPGNVDSSGIXHNIFN 2160
QY
2161 PPIIARYIRLHPHTHYSIR5TLEMLMGCDLNSC5MPLG5SKAISDAQITASSYFTNMFA 2220
Db
2161 PPIIARYIRLHPHTHYSIR5TLEMLMGCDLNSC5MPLG5SKAISDAQITASSYFTNMFA 2220
QY
2221 TWSPSKARLHLOGR5NANRPQVNNPKEMLOVDPOKTMKVGTGTGKSL5LTSYVYKBE 2280
Db
2221 TWSPSKARLHLOGR5NANRPQVNNPKEMLOVDPOKTMKVGTGTGKSL5LTSYVYKBE 2280
QY
2281 ISSSQDGHQWTLFFQNGKVKVFCQGNQDSFTPVNSLDPPLTRYLRIRHQP5WHQIALRM 2340
Db
2281 ISSSQDGHQWTLFFQNGKVKVFCQGNQDSFTPVNSLDPPLTRYLRIRHQP5WHQIALRM 2340
QY
2341 EVLGC6AODLY 2351
Db
2341 EVLGC6AODLY 2351

```

; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/205,226
; FILING DATE: 10-JUN-1988
; SEQ ID NO.2:
; LENGTH: 2351
; 517844-2

Query Match
Best Local Similarity 99.9%; Score 12404; DB 6; Length 2351;
Matches 2349; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 MOIELSTCFFCLLRFCFSATRRYYLGAVELSWDYMOSDLGSLPVDARPPRPVPSPPFN 60
DB 1 MOIELSTCFFCLLRFCFSATRRYYLGAVELSWDYMOSDLGSLPVDARPPRPVPSPPFN 60
QY 61 TSVVYKTLFVEFTDHLNIAKPRPPWGLLGTPTQAEVYDVIVVITLKNMASHPVSLHAV 120
DB 61 TSVVYKTLFVEFTDHLNIAKPRPPWGLLGTPTQAEVYDVIVVITLKNMASHPVSLHAV 120
QY 121 GSVYWKASGAEYDDOTSOREKEDDKVFPGGSHTVVQVLKXENGPMSADPLCLCTSYLSH 180
DB 121 GSVYWKASGAEYDDOTSOREKEDDKVFPGGSHTVVQVLKXENGPMSADPLCLCTSYLSH 180
QY 131 VDLVKDLSGLIGALLVCREGSLAKEKTQTLHKFILLFAVPDEGKSWHSETKNSLMQDRD 240
DB 131 VDLVKDLSGLIGALLVCREGSLAKEKTQTLHKFILLFAVPDEGKSWHSETKNSLMQDRD 240
QY 241 AASARAWPKMHTVNGVYVNSLPCLTGCHEKSVVWVHVI GMGTTPEVHSIFLECHTFLVRNH 300
DB 241 AASARAWPKMHTVNGVYVNSLPCLTGCHEKSVVWVHVI GMGTTPEVHSIFLECHTFLVRNH 300
QY 301 ROASLEISPIITFLTAQTLMLDGLQFLLSCHISSHODGMEAYVKYDSCPEBPOLRMKNE 360
DB 301 ROASLEISPIITFLTAQTLMLDGLQFLLSCHISSHODGMEAYVKYDSCPEBPOLRMKNE 360
QY 361 EADYDDDLTQSEMDVVRFPDDNSPSFIQIRSVAKKPKTWVHYIAABEEDWDYAPLVIA 420
DB 361 EADYDDDLTQSEMDVVRFPDDNSPSFIQIRSVAKKPKTWVHYIAABEEDWDYAPLVIA 420
QY 421 PDDRSKYQYLNNGPQRI GRKYKVRFWAYTDETFKTRAIQHESGILGPLLYGVGDTL 480
DB 421 PDDRSKYQYLNNGPQRI GRKYKVRFWAYTDETFKTRAIQHESGILGPLLYGVGDTL 480
QY 481 LIIFKNQASRPNIYPHGITDVRPLYSRLPKGVXHLKDPFPLPCEIIPKYKVTIVVEDGP 540
DB 481 LIIFKNQASRPNIYPHGITDVRPLYSRLPKGVXHLKDPFPLPCEIIPKYKVTIVVEDGP 540
QY 541 TKSDPRCLTRYSSVFNMERDLASGLIOPLLICYKESVDQRGNQIMSDKRNVLFSVFDE 600
DB 541 TKSDPRCLTRYSSVFNMERDLASGLIOPLLICYKESVDQRGNQIMSDKRNVLFSVFDE 600
QY 601 NRSWLTENIORELPNPAQVQLEDPEFOASNTMHSGVVDLSLQSVCLHEVAYWYILS 660
DB 601 NRSWLTENIORELPNPAQVQLEDPEFOASNTMHSGVVDLSLQSVCLHEVAYWYILS 660
QY 661 ICAQTDPLSVFSGYTFKHVMYEDTLTLPFPGSETVFMSENPGILWILGCHNSDFNRG 720
DB 661 ICAQTDPLSVFSGYTFKHVMYEDTLTLPFPGSETVFMSENPGILWILGCHNSDFNRG 720
QY 721 MTALLKVSCKDNTGDDYEDSDIYSAVLLSKNNAIEPRSPSONSRHPSSTROKQNATTI 780
DB 721 MTALLKVSCKDNTGDDYEDSDIYSAVLLSKNNAIEPRSPSONSRHPSSTROKQNATTI 780
QY 781 PENDIEKTDPPFAHRTPEKIQNVSSDDLMLLROSPTPHGLSLSDLOEAKVETFSDDPS 840
DB 781 PENDIEKTDPPFAHRTPEKIQNVSSDDLMLLROSPTPHGLSLSDLOEAKVETFSDDPS 840
QY 841 PCAIDSNLSLSEMTFRPQLHSHSGDMVTFPPESGLQRLNEKLGTTAAATLKKLDPKVSST 900
DB 841 PCAIDSNLSLSEMTFRPQLHSHSGDMVTFPPESGLQRLNEKLGTTAAATLKKLDPKVSST 900
QY 901 SNNLISSTIPSONLAAGTNTSSLGPPSPVHYVDSQDITTLFGKKSPLTSCGGPISLSEE 960
DB 901 SNNLISSTIPSONLAAGTNTSSLGPPSPVHYVDSQDITTLFGKKSPLTSCGGPISLSEE 960
QY 961 NNDKLLSGGLMNSQESSWGKNVSTESGRUFGKRAHGPALLITKDNALFKVYSISLLKTN 1020
DB 961 NNDKLLSGGLMNSQESSWGKNVSTESGRUFGKRAHGPALLITKDNALFKVYSISLLKTN 1020
QY 1021 KTSNNSATNRKTHIDGPSLJITENSPSVWONLLESDETEKKTPTPLIHDRMLKDNATLRL 1080
DB 1021 KTSNNSATNRKTHIDGPSLJITENSPSVWONLLESDETEKKTPTPLIHDRMLKDNATLRL 1080
QY 1081 NMSNKTSSKNMEMVQOKKGGPIPPDAQNPMSPFKMLFLPESARWIORCHGKNSLNG 1140
DB 1081 NMSNKTSSKNMEMVQOKKGGPIPPDAQNPMSPFKMLFLPESARWIORCHGKNSLNG 1140
QY 1141 QGSPFKQVSLGPEKSVGQNFLESEKNKVYVKGGEFTKDVGLKEMVPPSSRNLFUTNLND 1200
DB 1141 QGSPFKQVSLGPEKSVGQNFLESEKNKVYVKGGEFTKDVGLKEMVPPSSRNLFUTNLND 1200
QY 1201 LHENNTNHOEKI QEEIEKKETLLOENVVLPOIHTVGTGKNFMKNLFLLSTRQNVGSGYD 1260
DB 1201 LHENNTNHOEKI QEEIEKKETLLOENVVLPOIHTVGTGKNFMKNLFLLSTRQNVGSGYD 1260
QY 1261 GAYAPVLQDFRSLNDS TNRTKHTAHFSKGBEENLEGLGNQTKQIVKYYACTTISNT 1320
DB 1261 GAYAPVLQDFRSLNDS TNRTKHTAHFSKGBEENLEGLGNQTKQIVKYYACTTISNT 1320
QY 1321 SQQNFVTOQRKRAL KQRLPLEETELEKRIIVDDTSTOWSNKMKHLTPSTLTQIDYNEKE 1380
DB 1321 SQQNFVTOQRKRAL KQRLPLEETELEKRIIVDDTSTOWSNKMKHLTPSTLTQIDYNEKE 1380
QY 1381 KGATQSPSLPCLIRSHSIPOANRSPPLIAKVSPSPSIRPIYLTREVLPDQNSSHLPAASY 1440
DB 1381 KGATQSPSLPCLIRSHSIPOANRSPPLIAKVSPSPSIRPIYLTREVLPDQNSSHLPAASY 1440
QY 1441 RKXDSGVQESSHFIQGAKKNNLSLAILTLEMTGQOREVGSISLTSATNSVTYKVENTVLP 1500
DB 1441 RKXDSGVQESSHFIQGAKKNNLSLAILTLEMTGQOREVGSISLTSATNSVTYKVENTVLP 1500
QY 1501 KPDLPKTSKVELLPKVHIYQKLPFTBTSNGSPGHLDLVREGSLLOQTEGAIKWNEANRP 1560
DB 1501 KPDLPKTSKVELLPKVHIYQKLPFTBTSNGSPGHLDLVREGSLLOQTEGAIKWNEANRP 1560
QY 1561 GKVPFLRVATESSAKTSPKLDPLAWDNHYGTQIPKEWKSQESKSPKTAPEKKDITLSL 1620
DB 1561 GKVPFLRVATESSAKTSPKLDPLAWDNHYGTQIPKEWKSQESKSPKTAPEKKDITLSL 1620
QY 1621 NACESNEAAAINEGQNKPEIEVTWAKQGRTERLCSQNPVLRKHOREITRTTLQSDOEE 1680
DB 1621 NACESNEAAAINEGQNKPEIEVTWAKQGRTERLCSQNPVLRKHOREITRTTLQSDOEE 1680
QY 1681 IDYDDTISVENMKKEDFDIYDEDENQSPRSPOKKTGRHYPIAAVERLDWYMGSSSPHVLNR 1740
DB 1681 IDYDDTISVENMKKEDFDIYDEDENQSPRSPOKKTGRHYPIAAVERLDWYMGSSSPHVLNR 1740
QY 1741 AQSSVPOFKKVVQEFDTGSGFTQPLVGEVNEHLLGGLPVIRAEVDENIMVTFRNOASR 1800
DB 1741 AQSSVPOFKKVVQEFDTGSGFTQPLVGEVNEHLLGGLPVIRAEVDENIMVTFRNOASR 1800
QY 1801 PYSYSSLIISYEEEDQROGAEPKRNKVFKNETKTYFWKVQHMAFTKQEFDCWAYFSDV 1860
DB 1801 PYSYSSLIISYEEEDQROGAEPKRNKVFKNETKTYFWKVQHMAFTKQEFDCWAYFSDV 1860
QY 1861 DLEKDVHSGILGLLVCHTNTLNPAGROVTVQBFALFFITFDETKSWYFENNERCRA 1920
DB 1861 DLEKDVHSGILGLLVCHTNTLNPAGROVTVQBFALFFITFDETKSWYFENNERCRA 1920
QY 1921 PCNTQMEDPTFKENRFRHAINGYIMDTLPGJVAQDORIPWYLLSMGSNENIHHSFSGH 1980
DB 1921 PCNTQMEDPTFKENRFRHAINGYIMDTLPGJVAQDORIPWYLLSMGSNENIHHSFSGH 1980
QY 1981 VFTYRKKEEYKMALYNLYPGVFETVEMLPKAGIRWVECLIGEHLHAGMSTFLVYSNKC 2040
DB 1981 VFTYRKKEEYKMALYNLYPGVFETVEMLPKAGIRWVECLIGEHLHAGMSTFLVYSNKC 2040

```

QY 2041 QPLGASGHIRDFQITASQYQWAPKILARLHYSINAWSTKBPSPKVKYDLPAPMII 2100
Db 2041 QPLGASGHIRDFQITASQYQWAPKILARLHYSINAWSTKBPSPKVKYDLPAPMII 2100
QY 2101 HGKTCQARQKFSLSYISQFIIMYSLDGKKWQTYRGNSGTGLMVFFGNVDSSGKENIEN 2160
Db 2101 HGKTCQARQKFSLSYISQFIIMYSLDGKKWQTYRGNSGTGLMVFFGNVDSSGKENIEN 2160
QY 2161 PPIIABYIIRLPHPHYIRSLTLMELMGCDLNSCSMPLGMSKAIASDAQTASSYFTNMFA 2220
Db 2161 PPIIABYIIRLPHPHYIRSLTLMELMGCDLNSCSMPLGMSKAIASDAQTASSYFTNMFA 2220
QY 2221 TWSPSKARLHLQGRSNAWRPQVNNPKXEMLOVDFOKTMKVGTGTVQVKSLTSMYVKEFL 2280
Db 2221 TWSPSKARLHLQGRSNAWRPQVNNPKXEMLOVDFOKTMKVGTGTVQVKSLTSMYVKEFL 2280
QY 2281 ISSQDGHQWTLFFQNGKVKVFGQNDSPFPVNSLDPPLLTRYLZIHQFQSWVHQIALRM 2340
Db 2281 ISSQDGHQWTLFFQNGKVKVFGQNDSPFPVNSLDPPLLTRYLZIHQFQSWVHQIALRM 2340
QY 2341 EVLGCEAQDLY 2351
Db 2341 EVLGCEAQDLY 2351

RESULT 3
US-08-366-851A-2
; Sequence 2, Application US/08366851A
; Patent No. 5681746
; GENERAL INFORMATION:
; APPLICANT: Bodner, Mordechai
; APPLICANT: De Polo, Nicolas J.
; APPLICANT: Hsu, David Chi-Tang
; APPLICANT: Chang, Steven
; TITLE OF INVENTION: Retroviral Delivery of Full Length Factor VIII
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Viagene, Inc.
; STREET: 11055 Roselle Street
; CITY: San Diego
; STATE: California
; COUNTRY: U.S.A.
; ZIP: 92121
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/366,851A
; FILING DATE:
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Chambers, Daniel M.
; REGISTRATION NUMBER: 34,561
; REFERENCE/DOCKET NUMBER: 930049, 438
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 452-1288
; TELEFAX: (619) 452-2616
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2351 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-366-851A-2
Query Match: 99.9%; Score 12401; DB 1; Length 2351;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 2349; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 MQIELSTCFELCLLRCFSAATRRYYLGAVELSWDYNQSDLGELPVDARPPVPKSPFPN 60

Db 1 MQIELSTCFELCLLRCFSAATRRYYLGAVELSWDYNQSDLGELPVDARPPVPKSPFPN 60
QY 61 TSVYVKTLFVEFTDELFENIAKPRPPMGLGPTTQAEVYDVTWVITLKNWASHPVSLHAV 120
Db 61 TSVYVKTLFVEFTDELFENIAKPRPPMGLGPTTQAEVYDVTWVITLKNWASHPVSLHAV 120
QY 121 GYSYWKASBGAIDYDQTSOREKEDDKVPFGSGSHYVWVLKENGPMASDPLCLTYSYLSH 180
Db 121 GYSYWKASBGAIDYDQTSOREKEDDKVPFGSGSHYVWVLKENGPMASDPLCLTYSYLSH 180
QY 181 VDLVXLNSGLLGALLVCREGSLAKEKQTLHKFILLPAVDEGKSWHSETKNSLMQDSD 240
Db 181 VDLVXLNSGLLGALLVCREGSLAKEKQTLHKFILLPAVDEGKSWHSETKNSLMQDSD 240
QY 241 AASARAWPKHVTNGVYVNRSLPGLIGCHRKSVYVHVGMTTPEVHSIFLEGHTFLVRNH 300
Db 241 AASARAWPKHVTNGVYVNRSLPGLIGCHRKSVYVHVGMTTPEVHSIFLEGHTFLVRNH 300
QY 301 RCASLEISPIITELTAQTLMDLGOFLSCHISSHGDGMEAYKVYDSCPEEPQLMKONE 360
Db 301 RCASLEISPIITELTAQTLMDLGOFLSCHISSHGDGMEAYKVYDSCPEEPQLMKONE 360
QY 361 EAEDYDDDLTDSEMDVVRPDDDNPSFQIRSVAKKHPTWVHYTAABEEDWDYAPLVIA 420
Db 361 EAEDYDDDLTDSEMDVVRPDDDNPSFQIRSVAKKHPTWVHYTAABEEDWDYAPLVIA 420
QY 421 PDDRSYKQYLNNGPQRIGRKYKVRPMAYTDETFKTREAIQHEGILGLPILYGEVDTL 480
Db 421 PDDRSYKQYLNNGPQRIGRKYKVRPMAYTDETFKTREAIQHEGILGLPILYGEVDTL 480
QY 481 LIIFKNOASRPVNTYPHGITDVRPLYSRRLPGVXELKDFPILPGEIPKYKWTVTVEGP 540
Db 481 LIIFKNOASRPVNTYPHGITDVRPLYSRRLPGVXELKDFPILPGEIPKYKWTVTVEGP 540
QY 541 TXSDPRCLTRYYSFVNMERDLASGLIGPLICYKESVDQGNQIMSKRNVLFSVDE 600
Db 541 TXSDPRCLTRYYSFVNMERDLASGLIGPLICYKESVDQGNQIMSKRNVLFSVDE 600
QY 601 NRSWYLTENIQRPFLNPAGVQLEDPEFOASNTMHSINGVYVFDLSQISVCLHVAWYILS 660
Db 601 NRSWYLTENIQRPFLNPAGVQLEDPEFOASNTMHSINGVYVFDLSQISVCLHVAWYILS 660
QY 661 IGAOTDELVSVPESGYTFKHVYVDTLTLPPESGTVFMSMENPGLWILGCHNSDFRARG 720
Db 661 IGAOTDELVSVPESGYTFKHVYVDTLTLPPESGTVFMSMENPGLWILGCHNSDFRARG 720
QY 721 MTALLKVSDDXNTGDYVEDSYEDISAYLLSKNNAIEPRFSQNSRHPSTROKQFNATTI 780
Db 721 MTALLKVSDDXNTGDYVEDSYEDISAYLLSKNNAIEPRFSQNSRHPSTROKQFNATTI 780
QY 781 PENDIEKTPWFAHRTMPKIQNTSSSLLMLLRQSPHGLSLDLOEAKYETSDSDPS 840
Db 781 PENDIEKTPWFAHRTMPKIQNTSSSLLMLLRQSPHGLSLDLOEAKYETSDSDPS 840
QY 841 PGALDSNNLSLSEMTFRPQLHESGDMVTPPSGLQRLNEKLGTTAAATELKXLDPKVAST 900
Db 841 PGALDSNNLSLSEMTFRPQLHESGDMVTPPSGLQRLNEKLGTTAAATELKXLDPKVAST 900
QY 901 SNNLISTIPSONLAAGTNTSSLGPPSMFVYDSQDITLFGKKSPLTESGPPSLSEE 960
Db 901 SNNLISTIPSONLAAGTNTSSLGPPSMFVYDSQDITLFGKKSPLTESGPPSLSEE 960
QY 961 NNDKSLLESGLMNSOESSGKNVSTESGRLFKGRAGHPALLTKDNALFKVLSILLTN 1020
Db 961 NNDKSLLESGLMNSOESSGKNVSTESGRLFKGRAGHPALLTKDNALFKVLSILLTN 1020
QY 1021 KTSNNSATNRKTHJDGCPILLIENSVPVWNTLESDETEKVTPLTHDMLMDKNATLRL 1080
Db 1021 KTSNNSATNRKTHJDGCPILLIENSVPVWNTLESDETEKVTPLTHDMLMDKNATLRL 1080
QY 1081 NNSWNTTSSKNMVMVQKKEGPIPPDACNPMSFFKMLFLPESARWQRTGHGKNSLNSG 1140

Db 1081 NHMSNKTSSKNMVMQCKEGP1PPDAQNPMQSFPMFLPESARW1QRTKCNLSNG 1140
QY 1141 QGSPKQVLSLGPESKVBGONFLSEKXKVVGKGEPTKDVGLKEMVPPSSRNLFNTLND 1200
Db 1141 QGSPKQVLSLGPESKVBGONFLSEKXKVVGKGEPTKDVGLKEMVPPSSRNLFNTLND 1200
QY 1201 LHENNTNOKKIQEIEKKETLIQENVVLPGTHVTGTFKNFMKNLFLLLSTRQNVGSYD 1260
Db 1201 LHENNTNOKKIQEIEKKETLIQENVVLPGTHVTGTFKNFMKNLFLLLSTRQNVGSYD 1260
QY 1261 GAYAPVLDQFRSLNSTNTEKHTAHFSGKGBEENLEGNGOTKQIVEKYACTTRISPNT 1320
Db 1261 GAYAPVLDQFRSLNSTNTEKHTAHFSGKGBEENLEGNGOTKQIVEKYACTTRISPNT 1320
QY 1321 SQCNFVTRSKALKQFRPLBETELEK311VDDTSTOWSKXMKHLTPSTLTQIDVNEKE 1380
Db 1321 SQCNFVTRSKALKQFRPLBETELEK311VDDTSTOWSKXMKHLTPSTLTQIDVNEKE 1380
QY 1381 KGAITQSPUSDCLTRSHSIPQANRSPLP1AKVSSPPIRYLTVLFDQNSSHLPAA5Y 1440
Db 1381 KGAITQSPUSDCLTRSHSIPQANRSPLP1AKVSSPPIRYLTVLFDQNSSHLPAA5Y 1440
QY 1441 RKQSGVQESSHFLQAKKNNLSLAILLTXMGDQREVGSLGTSATNSVTYKKVNTVLP 1500
Db 1441 RKQSGVQESSHFLQAKKNNLSLAILLTXMGDQREVGSLGTSATNSVTYKKVNTVLP 1500
QY 1501 KPDLPKTSKVLLPKVHIYQKDLFPPTSTNSGSPGHLQTEGSLQTEGAIKWNEANRP 1560
Db 1501 KPDLPKTSKVLLPKVHIYQKDLFPPTSTNSGSPGHLQTEGSLQTEGAIKWNEANRP 1560
QY 1561 GKVPFLRVATRESAKTPSKLLPLAWDNHYGTQIPKEWKSGQEKPEKTAFFKKDITLSL 1620
Db 1561 GKVPFLRVATRESAKTPSKLLPLAWDNHYGTQIPKEWKSGQEKPEKTAFFKKDITLSL 1620
QY 1621 NACESNHAIAANEGONKPEIBVTWAKQSTERI CSQPPVLKRQRETRITLOSQEE 1680
Db 1621 NACESNHAIAANEGONKPEIBVTWAKQSTERICSONPPVLKRQRETRITLOSQEE 1680
QY 1681 IDYDITI SVEMKKEPDIYDENOSPRSFOKTRHYFIAAVERLWDYGMSSSPHVLNR 1740
Db 1681 IDYDITI SVEMKKEPDIYDENOSPRSFOKTRHYFIAAVERLWDYGMSSSPHVLNR 1740
QY 1741 AQSGVPOFKKVFOEFTDGSQTPLYRGELNEHLGLGPIYIRAEVEDNIMVTFNQASR 1800
Db 1741 AQSGVPOFKKVFOEFTDGSQTPLYRGELNEHLGLGPIYIRAEVEDNIMVTFNQASR 1800
QY 1801 PYSFYSSLSIYSEEDORQGAERKNEKNTKTYFWKHQHMAPT KDBPDCCKAWAYFSDV 1860
Db 1801 PYSFYSSLSIYSEEDORQGAERKNEKNTKTYFWKHQHMAPT KDBPDCCKAWAYFSDV 1860
QY 1861 DLEKDVHSLGLGPLVCHTNTLNPAGRQVTVQEFALPFTIPDET KSWYFTENMERNCR 1920
Db 1861 DLEKDVHSLGLGPLVCHTNTLNPAGRQVTVQEFALPFTIPDET KSWYFTENMERNCR 1920
QY 1921 PCNIOMEDPTKENYRFAINGYIMDTPLGLWAQDQIRWYLLSMGSGNENIHSHPGSH 1980
Db 1921 PCNIOMEDPTKENYRFAINGYIMDTPLGLWAQDQIRWYLLSMGSGNENIHSHPGSH 1980
QY 1981 VFTVRKKEEYKXALYNLPGVFETVEMLPKAGIWRVECLIGELHACGMSTLFLVYSNKC 2040
Db 1981 VFTVRKKEEYKXALYNLPGVFETVEMLPKAGIWRVECLIGELHACGMSTLFLVYSNKC 2040
QY 2041 QTPLGASGHIRDFOITASQGYQWAPKLARLHYSGSINAWSTKEPFSWIKVDLLAPMII 2100
Db 2041 QTPLGASGHIRDFOITASQGYQWAPKLARLHYSGSINAWSTKEPFSWIKVDLLAPMII 2100
QY 2101 HGIKTQGAHQESSLIYSQFIMYSLDGKKQCTYRGNSTGTLMFPFGVNDVSSGKHNI FN 2160
Db 2101 HGIKTQGAHQESSLIYSQFIMYSLDGKKQCTYRGNSTGTLMFPFGVNDVSSGKHNI FN 2160
QY 2161 PPIIARYIRLJPHTHYSIRSTIRMLMGLDNLSCMPLGNESKAI SDAQITASSYFTNMF 2220
Db 2161 PPIIARYIRLJPHTHYSIRSTIRMLMGLDNLSCMPLGNESKAI SDAQITASSYFTNMF 2220

QY 2221 TWSPSKAPLHIOGRENANRPQVNNPKWLQVDFOKTMKVGTGTTQGVKSLTSTMTVKBEFL 2280
Db 2221 TWSPSKAPLHIOGRENANRPQVNNPKWLQVDFOKTMKVGTGTTQGVKSLTSTMTVKBEFL 2280
QY 2281 1SSSODGQHWLTFQNGKVKVFOGQNDSTFTPVVNSLDPLLTTRYLRIHPQSVHQAIALRM 2340
Db 2281 1SSSODGQHWLTFQNGKVKVFOGQNDSTFTPVVNSLDPLLTTRYLRIHPQSVHQAIALRM 2340
QY 2341 EVLGCEAQDLY 2351
Db 2341 EVLGCEAQDLY 2351
RESULT 4
US-08-121-202-2
; Sequence 2, Application US/08121202
; Patent No. 5563045
; GENERAL INFORMATION:
; APPLICANT: Pittman, Debra
; APPLICANT: Rehenculla, Alhawaz
; APPLICANT: Wozney, Cohn M.
; APPLICANT: Kaufman, Randal J.
; TITLE OF INVENTION: CHIMERIC PROCOAGULANT PROTEINS
; NUMBER OF SEQUENCES: 27
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Legal Affairs, Genetics Institute, Inc.
; STREET: 87 Cambridgepark Drive
; CITY: Cambridge
; STATE: MA
; COUNTRY: USA
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/121,202
; FILING DATE: 14-SEP-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Weinert, M. C.
; REGISTRATION NUMBER: 31,544
; REFERENCE/DOCKET NUMBER: GI 5195A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 876-1210 X8574
; TELEFAX: (617) 876-5851
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2351 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-121-202-2
Query Match 99.8%; Score 12397; DB 1; Length 2351;
Best local Similarity 99.9%; Pred. No. 0;
Matches 2348; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
QY 1 MQEILSTCFCLLRCFCFSATRRYYLGAVELSWDYMOSDLGELPVDARFPFPRVKSPPFN 60
Db 1 MQEILSTCFCLLRCFCFSATRRYYLGAVELSWDYMOSDLGELPVDARFPFPRVKSPPFN 60
QY 61 TSVVYKXTLTFVEFTDHLFNIAKRPMPWMLLGPTIOAEVYDVTVVITLKNMASHPVSLHAV 120
Db 61 TSVVYKXTLTFVEFTDHLFNIAKRPMPWMLLGPTIOAEVYDVTVVITLKNMASHPVSLHAV 120
QY 121 GVSYWKASGAEYDDQTSQREKDDKVPQSGSHYVWQVLKENGPMASDPLCLTYSLSH 180
Db 121 GVSYWKASGAEYDDQTSQREKDDKVPQSGSHYVWQVLKENGPMASDPLCLTYSLSH 180
QY 181 VDLVKDLNSGLIGALLVCRBGSIAKEXTQTLHKFTILLFAVFDGKSWHSETNSLMQDRD 240

181 VDLVXDJNSGLI GALLVCRBGS LAKEKTQTLHKPILLFAVDEBGSXWSSTKNSLQCD3D 240
 241 AASARAPKXHTVNGVYNSLPGJGCHRSVYVHVHVGMTTFEVSIFLEHGTTFVRNH 300
 241 AASARAPKXHTVNGVYNSLPLGLICHRKSVYVHVHVGMTTFEVSIFLEHGTTFVRNE 300
 301 RQASLEISPIITFTAOQLMDLQGFJLSCHEISSHQHDMGEAYVKVDCBEPQLRMKNE 360
 301 RQASLEISPIITFTAOQLMDLQGFJLSCHEISSHQHDMGEAYVKVDCBEPQLRMKNE 360
 361 EADYDDDLTDSMDVVRPDDNSPFIQIRSVAKXHPKTVWHYIAAEEDMDYAPJVL4 420
 361 EADYDDDLTDSMDVVRPDDNSPFIQIRSVAKXHPKTVWHYIAAEEDMDYAPJVL4 420
 421 PDRSRYSQVQLNNGPQIRGRKYKVRFMAYTDETEKTRAIQHESGILGPLLXGVEGDTL 480
 421 PDRSRYSQVQLNNGPQIRGRKYKVRFMAYTDETEKTRAIQHESGILGPLLXGVEGDTL 480
 481 LIIFKQAGRPYNIYPEG-TDVARJXSRRLPKGVGHLKDPPIIPGSIFFKYKWTVTVEDGP 540
 481 LIIFKQAGRPYNIYPEG-TDVARJXSRRLPKGVGHLKDPPIIPGSIFFKYKWTVTVEDGP 540
 541 TKSDDPCLTRYYSFVNMRDLASGLIGPLLIICYESVDORGNOIMSDKENVILPSVFE 600
 541 TKSDDPCLTRYYSFVNMRDLASGLIGPLLIICYESVDORGNOIMSDKENVILPSVFE 600
 601 NRSWYLTENIQRLPNPAGVOLBDDPFOASNIHWSINGVYVDSLOJLVCLHEVAYVWILS 660
 601 NRSWYLTENIQRLPNPAGVOLBDDPFOASNIHWSINGVYVDSLOJLVCLHEVAYVWILS 660
 661 IGAQTDPLSVFSGYTFKHMVYEDTLTPFPGSETVFMSENPGWLILGCHNSDFRNG 720
 661 IGAQTDPLSVFSGYTFKHMVYEDTLTPFPGSETVFMSENPGWLILGCHNSDFRNG 720
 721 MTALLKVVSCDKNTGYEDSYDZIDJAYLLSKNAIEPRSPNSRHPSTROQFNATTI 780
 721 MTALLKVVSCDKNTGYEDSYDZIDJAYLLSKNAIEPRSPNSRHPSTROQFNATTI 780
 781 PENDIEKTPWFARHTPMKPIQNVSSDLMILRQGTPTGHLGSLSLDQAKYFTFSDDP8 840
 781 PENDIEKTPWFARHTPMKPIQNVSSDLMILRQGTPTGHLGSLSLDQAKYFTFSDDP8 840
 841 PGAIIDNNLSEMTHTPQLHSGDMVFTPESGLOLRNEKLGTTAATLKLDFKVSST 900
 841 PGAIIDNNLSEMTHTPQLHSGDMVFTPESGLOLRNEKLGTTAATLKLDFKVSST 900
 901 SNNLISTIPSDNLAAGTDNTSSLGPPMPVHYDSQIDTTLFGKSSPLTESGGLSLSEE 960
 901 SNNLISTIPSDNLAAGTDNTSSLGPPMPVHYDSQIDTTLFGKSSPLTESGGLSLSEE 960
 961 NNDKSLJESGLMNSQESSWGKNVSSSTESGRLFKGKAHGPALLTKDNAJFKVSI5LLKTN 1020
 961 NNDKSLJESGLMNSQESSWGKNVSSSTESGRLFKGKAHGPALLTKDNAJFKVSI5LLKTN 1020
 1021 KYSNNATNRKTHIDGSPILLIENSPVQWQNILESDFEKKVPTLIHDMMLDKNATLRL 1080
 1021 KYSNNATNRKTHIDGSPILLIENSPVQWQNILESDFEKKVPTLIHDMMLDKNATLRL 1080
 1081 NWSNKTTSKXNEMVQCKEGPIPDQAQNPDMSPFKYLFPLPESAEWIORTHGKNSLNSG 1140
 1081 NWSNKTTSKXNEMVQCKEGPIPDQAQNPDMSPFKYLFPLPESAEWIORTHGKNSLNSG 1140
 1141 QGSPKQLVSLGPEKSVGEGNFJSEKNKVVVGKGETKDVGLKEMVFPSPSRNLFJTNLDN 1200
 1141 QGSPKQLVSLGPEKSVGEGNFJSEKNKVVVGKGETKDVGLKEMVFPSPSRNLFJTNLDN 1200
 1201 LHENNTNHEKKIQEIEIEKKEKTIQENVLPOHTVTGKQNFKNLFLJLSTRQWEGSYD 1260
 1201 LHENNTNHEKKIQEIEIEKKEKTIQENVLPOHTVTGKQNFKNLFLJLSTRQWEGSYD 1260
 1261 GAVAPVLOQDFRSLNDSTNRKKTATAHFSKKGEEENLEGLGNQKQIVEXYACTTRISPNT 1320

1261 GAVAPVLOQDFRSLNDSTNRKKTATAHFSKKGEEENLEGLGNQKQIVEXYACTTRISPNT 1320
 1321 SQONFVTOBSKALQKFRPLPIEBTELEKRIIVDDTSTOWSKMKHLTPSTLLQIDYNEKE 1380
 1321 SQONFVTOBSKALQKFRPLPIEBTELEKRIIVDDTSTOWSKMKHLTPSTLLQIDYNEKE 1380
 1381 KGAITQSPSLDCLTRSHSIPQANRSPPIAKVSSPPSIRPIYLTVLFOONSHSLPAA5Y 1440
 1381 KGAITQSPSLDCLTRSHSIPQANRSPPIAKVSSPPSIRPIYLTVLFOONSHSLPAA5Y 1440
 1441 RKDGSVQSSSHFLQGAKKNNLSLAILTLEMTGDQREVGLSGTSAATNSVYTKKVENTVLP 1500
 1441 RKDGSVQSSSHFLQGAKKNNLSLAILTLEMTGDQREVGLSGTSAATNSVYTKKVENTVLP 1500
 1501 KPDLPKTSQKVELLPKVHIYQKDLPP7ETNSGSPGHLDLVBGSLLOCTGEGAKWNEANRP 1560
 1501 KPDLPKTSQKVELLPKVHIYQKDLPP7ETNSGSPGHLDLVBGSLLOCTGEGAKWNEANRP 1560
 1561 GKVPFLRVATESAKTPSKLLDPLAWDNHYGTQIPKEEMKSOBKPEKTAFAKKOTILSL 1620
 1561 GKVPFLRVATESAKTPSKLLDPLAWDNHYGTQIPKEEMKSOBKPEKTAFAKKOTILSL 1620
 1621 NACESNHAIAANEGONKPEIEVTVAKQGRTERLCSQNPVILKRHQREITRTTLOSQDEE 1680
 1621 NACESNHAIAANEGONKPEIEVTVAKQGRTERLCSQNPVILKRHQREITRTTLOSQDEE 1680
 1681 IDYDDTISVEMKKEPDIYDEDEBENOSPRSFOKTRHYFFIAAVERLWDYGNSSPHVLRN 1740
 1681 IDYDDTISVEMKKEPDIYDEDEBENOSPRSFOKTRHYFFIAAVERLWDYGNSSPHVLRN 1740
 1741 AQSGSVQPKVKKVVOFBEFTDGSFTQPLRGELNEHLGLGPIRAEVEDNIMVTFRNOASR 1800
 1741 AQSGSVQPKVKKVVOFBEFTDGSFTQPLRGELNEHLGLGPIRAEVEDNIMVTFRNOASR 1800
 1801 PYSFVSSLIISYEDORQGAEPKRNFKVNETKTYFKVOHHVAPKYDEDFCKAWAYFSDV 1860
 1801 PYSFVSSLIISYEDORQGAEPKRNFKVNETKTYFKVOHHVAPKYDEDFCKAWAYFSDV 1860
 1861 DLEKDVHSLIGLPLVCHTNTLNPAHGROVTVQEFALPFIETETKSWYFETENMERNCA 1920
 1861 DLEKDVHSLIGLPLVCHTNTLNPAHGROVTVQEFALPFIETETKSWYFETENMERNCA 1920
 1921 PCNIQMEDPTFKENYRFAHNGYIMDTLPLGLVMAQDQIRWYLLSNGSNENIHSIEFSGH 1980
 1921 PCNIQMEDPTFKENYRFAHNGYIMDTLPLGLVMAQDQIRWYLLSNGSNENIHSIEFSGH 1980
 1981 VFTVRKBEYKVALNLYPGVETVEMLPFKAGIWRVECLIGEHLHAGMSTLFLVYSNKC 2040
 1981 VFTVRKBEYKVALNLYPGVETVEMLPFKAGIWRVECLIGEHLHAGMSTLFLVYSNKC 2040
 2041 QTPLGWASGHIRDPQITASGOYQWAPKLARLHYSGSINAWSTKSPFSWKVDLLAPMII 2100
 2041 QTPLGWASGHIRDPQITASGOYQWAPKLARLHYSGSINAWSTKSPFSWKVDLLAPMII 2100
 2101 HGKTCGARQKPSYIYSQFIIMYSJGKKWOTYRGNSTGTLMVFFGNVDSGGIKHNIIPN 2160
 2101 HGKTCGARQKPSYIYSQFIIMYSJGKKWOTYRGNSTGTLMVFFGNVDSGGIKHNIIPN 2160
 2161 PPIIARYIRLHPHTHYSIRSTLMEJWGCDLNSCMLGSKALISDAQITASSYTNMFA 2220
 2161 PPIIARYIRLHPHTHYSIRSTLMEJWGCDLNSCMLGSKALISDAQITASSYTNMFA 2220
 2221 TWSPSKARLHLQGRSNANRPVNNKEWLQVDFOKTMKVGTVTTCVKSLLTSMVVKFEL 2280
 2221 TWSPSKARLHLQGRSNANRPVNNKEWLQVDFOKTMKVGTVTTCVKSLLTSMVVKFEL 2280
 2281 ISSSQDQCHQWTLFFQNGKVKVFGQNDSTFPVNSLDPLLTTRYLRHHPQSWVHQIALRM 2340
 2281 ISSSQDQCHQWTLFFQNGKVKVFGQNDSTFPVNSLDPLLTTRYLRHHPQSWVHQIALRM 2340
 2341 EYJGCEAQDLY 2351
 2341 EYJGCEAQDLY 2351

RESULT 5
5422260-1
; Patent No. 5422260
; APPLICANT: KAUFMAN, RANDAL J.;PITTMAN, DEBRA D.;TOOLE,JOHN J.
; TITLE OF INVENTION: HUMAN FACTOR VIII: C MUTAINS
; NUMBER OF SEQUENCES: 15
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/883,936
; FILING DATE: 15-MAY-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 279,485
; FILING DATE: 02-DEC-1988; 09-DEC-1986
; APPLICATION NUMBER: 939,658
; FILING DATE: 09-DEC-1986
; APPLICATION NUMBER: 932,767
; FILING DATE: 18-NOV-1986
; APPLICATION NUMBER: 868,410
; FILING DATE: 29-MAY-1986
; SEQ ID NO:1:
; LENGTH: 2351
5422260-1

Query Match 99.8%; Score 12388; DB 6; Length 2351;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 2347; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 MCELSCTCFLLCLLPCEFSATRRYVYGAVELSDWYMQSDLGELPVDARPPRVPKSFFPN 60
DB 1 MCELSCTCFLLCLLPCEFSATRRYVYGAVELSDWYMQSDLGELPVDARPPRVPKSFFPN 60

QY 61 TSVVYKTLFVEFTDHLFNIAKPPMPWGLGPTQIARVYTVITLKNMASHPVSLHAV 120
DB 61 TSVVYKTLFVEFTVHLFNIAKPPMPWGLGPTQIARVYTVITLKNMASHPVSLHAV 120

QY 121 QVSVWKASEGABYDDQTSQRKEDDKVPFGSGHTYVVOVLKENGPMASDPLCLTVSYLSH 180
DB 121 QVSVWKASEGABYDDQTSQRKEDDKVPFGSGHTYVVOVLKENGPMASDPLCLTVSYLSH 180

QY 181 VDLKDLNSGLIGALLVCRGSLAKEKCTQTLHKFTLLFAVEDEGKSWSETKNSLMQDRD 240
DB 181 VDLKDLNSGLIGALLVCRGSLAKEKCTQTLHKFTLLFAVEDEGKSWSETKNSLMQDRD 240

QY 241 AASARAWPKMTVNGYVNRSLPGLIGCHRKSVYVHVICMGTTPEVHSIFLRGHTFLVANH 300
DB 241 AASARAWPKMTVNGYVNRSLPGLIGCHRKSVYVHVICMGTTPEVHSIFLRGHTFLVANH 300

QY 301 QASLEISPIITFLTAQTLMDLGOFLLECHISSHOHDGMEAYVKVDSCEPEPQLRMKNNE 360
DB 301 QASLEISPIITFLTAQTLMDLGOFLLECHISSHOHDGMEAYVKVDSCEPEPQLRMKNNE 360

QY 361 EARDYDDDLTDEMDVVRFDNDNSPFTQIRSVAKKPKTWVHYIAAEEEDWDVAPLVLA 420
DB 361 EARDYDDDLTDEMDVVRFDNDNSPFTQIRSVAKKPKTWVHYIAAEEEDWDVAPLVLA 420

QY 421 PDDRSYKSOYLNGPQRIGRKYKVRFWAYTDEFTKTEAIOHESGILGPLLYGEVGDTL 480
DB 421 PDDRSYKSOYLNGPQRIGRKYKVRFWAYTDEFTKTEAIOHESGILGPLLYGEVGDTL 480

QY 481 LIIFKNQASRPYNIYPHGITDVRPLYSRRLPKGVKHLKDFILPGEIPKYKWTVTVEDGP 540
DB 481 LIIFKNQASRPYNIYPHGITDVRPLYSRRLPKGVKHLKDFILPGEIPKYKWTVTVEDGP 540

QY 541 TKSDPRCLTRYYSFVNMRDLASGLIGPLIICYKESVDQRGNQIMSKRVNVLFSVPDE 600
DB 541 TKSDPRCLTRYYSFVNMRDLASGLIGPLIICYKESVDQRGNQIMSKRVNVLFSVPDE 600

QY 601 NRSWYLTEINIQRFLPNPAGVQLEDPEFOASNMHSINGYVFDLSQLSVCLHEVAYWYLLS 660
DB 601 NRSWYLTEINIQRFLPNPAGVQLEDPEFOASNMHSINGYVFDLSQLSVCLHEVAYWYLLS 660

QY 661 IGAQTDFLSVFSFGYTFKHKVYEDTLTLFPFSGETVFMSENFGILGCHNSDFNRNG 720

DB 661 IGAQTDFLSVFSFGYTFKHKVYEDTLTLFPFSGETVFMSENFGILGCHNSDFNRNG 720

QY 721 MTALLKVSSCDKNTGDDYEDSYEDISAYLLSKNNAIEPRFSQNSRHPSTKQKFNATTI 780
DB 721 MTALLKVSSCDKNTGDDYEDSYEDISAYLLSKNNAIEPRFSQNSRHPSTKQKFNATTI 780

QY 781 PENDIEKTDPPFAHRTMPKIQNVSSDLMLLQSPPTPHGLSLSDLOEAYETFSDDPS 840
DB 781 PENDIEKTDPPFAHRTMPKIQNVSSDLMLLQSPPTPHGLSLSDLOEAYETFSDDPS 840

QY 841 PGALDSNNSISEMTHFRPOLHSGDMVFTPSGQLRLNEKLGTTAAETLKKLDFKVSST 900
DB 841 PGALDSNNSISEMTHFRPOLHSGDMVFTPSGQLRLNEKLGTTAAETLKKLDFKVSST 900

QY 901 SNNLIITIPSDNLAAGTNTSSLGPPMPVHYDSQDQDTTLFGKXSSPLTSGGGLSSEE 960
DB 901 SNNLIITIPSDNLAAGTNTSSLGPPMPVHYDSQDQDTTLFGKXSSPLTSGGGLSSEE 960

QY 961 NNDSKLJESGLMNSQESSWGKNSVSTESGRLPFKGRAHGPALLTKDNALFKVSISSLKTN 1020
DB 961 NNDSKLJESGLMNSQESSWGKNSVSTESGRLPFKGRAHGPALLTKDNALFKVSISSLKTN 1020

QY 1021 KTSNNSATNRKTHIDGSPSLLIENSVPWQNILEDSTBPKVTPLIHDMMLDKNATALRL 1080
DB 1021 KTSNNSATNRKTHIDGSPSLLIENSVPWQNILEDSTBPKVTPLIHDMMLDKNATALRL 1080

QY 1081 NHMKNKTISSKNMNVQCKKEGPIPPDAQNPDMSFFKMLFLPESARWIQRTGHKNSLNSG 1140
DB 1081 NHMKNKTISSKNMNVQCKKEGPIPPDAQNPDMSFFKMLFLPESARWIQRTGHKNSLNSG 1140

QY 1141 QGPSPKOLVSLGPKSVGEGQNFLEKKNVVGKGETKQVGLKEMVFPSSSNLPLTNLDN 1200
DB 1141 QGPSPKOLVSLGPKSVGEGQNFLEKKNVVGKGETKQVGLKEMVFPSSSNLPLTNLDN 1200

QY 1201 LHENNTNHOEKKIQEEIEKKETLLIQENVLPQIHVTGTGKTNFMKNLFLLSSTQNVGEGSYD 1260
DB 1201 LHENNTNHOEKKIQEEIEKKETLLIQENVLPQIHVTGTGKTNFMKNLFLLSSTQNVGEGSYE 1260

QY 1261 GAYAPVLQDPRSLNDSNTRTKKHTAHFSSKKEEENLEGLGNQTIQIVEKVIACCTRI SPNT 1320
DB 1261 GAYAPVLQDPRSLNDSNTRTKKHTAHFSSKKEEENLEGLGNQTIQIVEKVIACCTRI SPNT 1320

QY 1321 SQQNFVTOBSKRALQKPLPLEETELKRIIVDDTSTQWSKNMKHLTPSTLTQIDYKEKE 1380
DB 1321 SQQNFVTOBSKRALQKPLPLEETELKRIIVDDTSTQWSKNMKHLTPSTLTQIDYKEKE 1380

QY 1381 KGATITQSPSLDCLPRSHSIPOANRSPLPIAKVSSFPPIRPIYLTTRVLFDQNSSHLPAASY 1440
DB 1381 KGATITQSPSLDCLPRSHSIPOANRSPLPIAKVSSFPPIRPIYLTTRVLFDQNSSHLPAASY 1440

QY 1441 RKXDSGVQESSHFLQGAKNKNSLAAILTLEMTGQREVGSLGTSAITNSVTYKKVNTVLP 1500
DB 1441 RKXDSGVQESSHFLQGAKNKNSLAAILTLEMTGQREVGSLGTSAITNSVTYKKVNTVLP 1500

QY 1501 KPDLPKTSGKVELLPKHVYIKOLFPPTETNSGSPCHLDLVEGSLLOGTEGAIKWNEANRP 1560
DB 1501 KPDLPKTSGKVELLPKHVYIKOLFPPTETNSGSPCHLDLVEGSLLOGTEGAIKWNEANRP 1560

QY 1561 GKVPFLVATESSAKTSPSKLLDPLAWDNHYGTQIPKEEWSQEKSPKTAFFKKDITLSL 1620
DB 1561 GKVPFLVATESSAKTSPSKLLDPLAWDNHYGTQIPKEEWSQEKSPKTAFFKKDITLSL 1620

QY 1621 NACSNHAIAINEGONKPEIEVTWAKQGTRELCSQNPVPLKHHQREIIRTTLOSQRE 1680
DB 1621 NACSNHAIAINEGONKPEIEVTWAKQGTRELCSQNPVPLKHHQREIIRTTLOSQRE 1680

QY 1681 IDYDDTISVEMKKEDFDIYDEDENOSPRSFKKTRHYFIAAVERLWDYGVSSSPHVLRNR 1740
DB 1681 IDYDDTISVEMKKEDFDIYDEDENOSPRSFKKTRHYFIAAVERLWDYGVSSSPHVLRNR 1740

QY 1741 AQOSVQPKKKVPQEBFTDGSFTQPLYRGLNEHLGLLGPYIAEVEDNINVTFRNQASR 1800

```

1741 AOSGSVPQFKKVFQBFQSDFTQPLRYGELNEHGLGLGYRAEVEDNIMVTFRNQASR 1800
1801 PYSFSSLSYVEDQDQGGAPRKNFKPNETKYFHWKVQHMAPTQDQDCKAWAYFSDV 1860
1801 PYSFSSLSYVEDQDQGGAPRKNFKPNETKYFHWKVQHMAPTQDQDCKAWAYFSDV 1860
1861 DLEKDVHSLGLGLLCHTNTLPAHGRQVTVQEPALFFTFIDETKSWYFTENNERNCRA 1920
1861 DLEKDVHSLGLGLLCHTNTLPAHGRQVTVQEPALFFTFIDETKSWYFTENNERNCRA 1920
1921 PCNIQMEDPTFKENYAFHAINGIMDTLPCLVMAQDQRIWYLLSGNSNENIHSIHFSGH 1980
1921 PCNIQMEDPTFKENYAFHAINGIMDTLPCLVMAQDQRIWYLLSGNSNENIHSIHFSGH 1980
1981 VFTVRKKEEYKALYNLYPGVFETVEMLPKSKAGIWRVECLIGSHLHAGMSTLFLVYSKNC 2040
1981 VFTVRKKEEYKALYNLYPGVFETVEMLPKSKAGIWRVECLIGSHLHAGMSTLFLVYSKNC 2040
2041 QTPPLGMSGHIRDFQITASQYQGWAPKLAHLHYSGSINAWSTKEPFSWKVDLLAPMII 2100
2041 QTPPLGMSGHIRDFQITASQYQGWAPKLAHLHYSGSINAWSTKEPFSWKVDLLAPMII 2100
2101 HGKKTQAKQKSSLYISOFIIMYSLDGKKWQYIRONSSTGTLNVFGNVDSSGIKENIEN 2160
2101 HGKKTQAKQKSSLYISOFIIMYSLDGKKWQYIRONSSTGTLNVFGNVDSSGIKENIEN 2160
2161 PPIIARYIRLHPHYGIRSTLRMELMGCDLNSCMPLGMESKAIQITASSYFTNVEA 2220
2161 PPIIARYIRLHPHYGIRSTLRMELMGCDLNSCMPLGMESKAIQITASSYFTNVEA 2220
2221 TWSPSKARLHLQGRSNARPOVNNPKELQVDFQKTKMKTGVTTOGVKSLLTSMVYKEFL 2280
2221 TWSPSKARLHLQGRSNARPOVNNPKELQVDFQKTKMKTGVTTOGVKSLLTSMVYKEFL 2280
2281 ISSQDGHQWTLFFQNGKVKVQGNODSTFVNVNSLDPPLLRILYRHPQSVHQIALRM 2340
2281 ISSQDGHQWTLFFQNGKVKVQGNODSTFVNVNSLDPPLLRILYRHPQSVHQIALRM 2340
2341 EVLGCEAQDLY 2351
2341 EVLGCEAQDLY 2351

```

RESULT 6

```

US-07-864-0045-4
; Sequence 4, Application US/07864004E
; Patent No. 5364771
; GENERAL INFORMATION:
; APPLICANT: Lollar, John S.
; TITLE OF INVENTION: Hybrid Human/Porcine Factor VIII
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Kilpatrick & Cody
; STREET: 1100 Peachtree Street
; CITY: Atlanta
; STATE: Georgia
; COUNTRY: US
; ZIP: 30309
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/864,004B
; FILING DATE: 07 APRIL 1992
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Pabst, Patrea L.
; REGISTRATION NUMBER: 31,284
; REFERENCE/DOCKET NUMBER: EMU106
; TELECOMMUNICATION INFORMATION:

```

```

; TELEPHONE: 404-815-6508
; TELEFAX: 404-815-6555
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2332 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHEetical: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: N-terminal
; ORIGINAL SOURCE:
; ORGANISM: Homo sapien
; TISSUE TYPE: Liver cDNA sequence
; US-07-864-0045-4

Query Match 99.0%; Score 12293; DB 1; Length 2332;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 2329; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 20 ATRRYYLGAVELSWDYMSDGLGELPVDARPPPRVPSKSPFNTSVVYKTLFVEFTDHPN 79
Db 1 ATRRYYLGAVELSWDYMSDGLGELPVDARPPPRVPSKSPFNTSVVYKTLFVEFTDHPN 60
QY 80 IAKPRPPWMLLGPITQAEVYDVVITLKNMASHPVSLHAGVSYWKASEGAEYDDQSQ 139
Db 61 IAKPRPPWMLLGPITQAEVYDVVITLKNMASHPVSLHAGVSYWKASEGAEYDDQSQ 120
QY 140 REKEDDKVPFGGSHYVMQVLKENGFMASDPLCLTYSYLSHVDLVKDLNSGLIGALLVCR 199
Db 121 REKEDDKVPFGGSHYVMQVLKENGFMASDPLCLTYSYLSHVDLVKDLNSGLIGALLVCR 180
QY 200 EGS LAKEKTQTLHKZILLPAVDEKSKWSHSETKNSLMQDRDAASARAPMHTVNGYVR 259
Db 181 EGS LAKEKTQTLHKZILLPAVDEKSKWSHSETKNSLMQDRDAASARAPMHTVNGYVR 240
QY 260 SLPLGLIGCHRSKVWHVITGMGTTPVHSITFLGHTPLVNRHQASLEISPIFLTAQTLL 319
Db 241 SLPLGLIGCHRSKVWHVITGMGTTPVHSITFLGHTPLVNRHQASLEISPIFLTAQTLL 300
QY 320 MDLQGFLLSCHISSHQHDMGEAYVVDVSCPEEPQRLMKNNSEAEYDDDLTSEMDEVVRP 379
Db 301 MDLQGFLLSCHISSHQHDMGEAYVVDVSCPEEPQRLMKNNSEAEYDDDLTSEMDEVVRP 360
QY 380 DDNSPSFIQIRSVAKKHPKTVWHYIAAEEDWDYAPLADDDRSYKSOYLNNGPQIG 439
Db 361 DDNSPSFIQIRSVAKKHPKTVWHYIAAEEDWDYAPLADDDRSYKSOYLNNGPQIG 420
QY 440 RYKVKVRFWAYTDETFKTRATQHESGILGPLLYGEVGTLLIIFKNOASRPVNTYPHGI 499
Db 421 RYKVKVRFWAYTDETFKTRATQHESGILGPLLYGEVGTLLIIFKNOASRPVNTYPHGI 480
QY 500 TDVRPLYGRRLPKGVKHLKOPILPGEIFKYKWTVTVEDGPKSDPCRLTRYSSFVNM 559
Db 481 TDVRPLYGRRLPKGVKHLKOPILPGEIFKYKWTVTVEDGPKSDPCRLTRYSSFVNM 540
QY 560 RDLASGLIGPLLI CYKESVDQGNQIMSDKRNVLFSVFDENRSWYLTENTQRFPLNPA 619
Db 541 RDLASGLIGPLLI CYKESVDQGNQIMSDKRNVLFSVFDENRSWYLTENTQRFPLNPA 600
QY 620 VQLEDPEFQASINHSINGYVFDLSQLSVCLHEVAYWYILSIGAOTDPLSVFFSGYTFKH 679
Db 601 VQLEDPEFQASINHSINGYVFDLSQLSVCLHEVAYWYILSIGAOTDPLSVFFSGYTFKH 660
QY 680 KXVYEDTTLTLPFSGETVFMGMENPGLMILGCHNSDFNRGMNTALLKVSXCDKNTGYE 739
Db 661 KXVYEDTTLTLPFSGETVFMGMENPGLMILGCHNSDFNRGMNTALLKVSXCDKNTGYE 720
QY 740 DSYEDISAYLLSKNNALIPRFSQNSRHPSTROKOFNATTIPENDIEKTDWFAHRTMP 799
Db 721 DSYEDISAYLLSKNNALIPRFSQNSRHPSTROKOFNATTIPENDIEKTDWFAHRTMP 780

```

QY 800 KIQNVSSDMLLRQSPHGLSLSDLOEAKYETSPDDPSCAIDSNNSLSGEMTHFRPQ 859
DB 781 KIQNVSSDMLLRQSPHGLSLSDLOEAKYETSPDDPSCAIDSNNSLSGEMTHFRPQ 840
QY 860 LHSQDMVFTPSGLOLRNEKLGITTAATELKKLDFKVSSTNNLITSTPSNLAAGTGN 919
DB 841 LHSQDMVFTPSGLOLRNEKLGITTAATELKKLDFKVSSTNNLITSTPSNLAAGTGN 900
QY 920 TSSLPPSPMVEHYDSOLDTTLFGKSSPLTSGGGLSLSEKNDSKLLBSGLNNSQESSW 979
DB 901 TSSLPPSPMVEHYDSOLDTTLFGKSSPLTSGGGLSLSEKNDSKLLBSGLNNSQESSW 960
QY 980 GKNVSTESGRLEKGRAGHPALLTKDNALFKVSISSLLKTKNTSNNNSATNRKTHIDGPSL 1039
DB 961 GKNVSTESGRLEKGRAGHPALLTKDNALFKVSISSLLKTKNTSNNNSATNRKTHIDGPSL 1020
QY 1040 LIENSPSVWONTLESDTEFKKVTPLIHDRMLDKNATLRLNHNMSNKTSSKNMVMVQOK 1099
DB 1021 LIENSPSVWONTLESDTEFKKVTPLIHDRMLDKNATLRLNHNMSNKTSSKNMVMVQOK 1080
QY 1100 KEGPIPPDAQNEDMFPFKMLFLPEGARWIQTHGKNSLNSGOGSPKQVLVSLGPEKSVBG 1159
DB 1081 KEGPIPPDAQNEDMFPFKMLFLPEGARWIQTHGKNSLNSGOGSPKQVLVSLGPEKSVBG 1140
QY 1160 QNFLSEKXKVVVGKGEFTKDVGLKEMVFPSSRNLFJTNLDNLHNNTHNQEKKIQEEIEK 1219
DB 1141 QNFLSEKXKVVVGKGEFTKDVGLKEMVFPSSRNLFJTNLDNLHNNTHNQEKKIQEEIEK 1200
QY 1220 KETLLOENVVLPOIHTVTGKFMKNLFLSTRQNVGSGYDAGAYAVIODFRSLNDSNTR 1279
DB 1201 KETLLOENVVLPOIHTVTGKFMKNLFLSTRQNVGSGYDAGAYAVIODFRSLNDSNTR 1260
QY 1280 TKKTAHFSKKEEENLEGLNQTQIWEKYACTTRISNTSQONFVTOORSK3ALQKPEL 1339
DB 1261 TKKTAHFSKKEEENLEGLNQTQIWEKYACTTRISNTSQONFVTOORSK3ALQKPEL 1320
QY 1340 PLEETELEKRIIIVDTSTOWSNMKGHTPSTUTQIDYNEKEKGAITQSPSLDCLTRSHSI 1399
DB 1321 PLEETELEKRIIIVDTSTOWSNMKGHTPSTUTQIDYNEKEKGAITQSPSLDCLTRSHSI 1380
QY 1400 POANSPPLIAKVSFSPRIPLTRVLFDQNSHPLPAASYEKDQSGVQESSHFLOGAKK 1459
DB 1381 POANSPPLIAKVSFSPRIPLTRVLFDQNSHPLPAASYEKDQSGVQESSHFLOGAKK 1440
QY 1460 NNL5ALIULTLEMTGQOREVSGSLGTSATNSVTKVENTVLPKPDLPTKSGKVELLPKVHI 1519
DB 1441 NNL5ALIULTLEMTGQOREVSGSLGTSATNSVTKVENTVLPKPDLPTKSGKVELLPKVHI 1500
QY 1520 YOKDLFPPTETSNNGSPGHLDLVBGSLLOQTEGAIKWNEANRPKVPLRVATESSAKTFSK 1579
DB 1501 YOKDLFPPTETSNNGSPGHLDLVBGSLLOQTEGAIKWNEANRPKVPLRVATESSAKTFSK 1560
QY 1580 LLDPLAWNHYGTQIPKEWKSQESKPEKTAPEKXDTILSLNACSNHAIAINSGQNK 1639
DB 1561 LLDPLAWNHYGTQIPKEWKSQESKPEKTAPEKXDTILSLNACSNHAIAINSGQNK 1620
QY 1640 ELEVTAQGRTERLCSQNPVLRKHOREITRTTLQSQDEEIDYDDTISVEMKEDFDIY 1699
DB 1621 ELEVTAQGRTERLCSQNPVLRKHOREITRTTLQSQDEEIDYDDTISVEMKEDFDIY 1680
QY 1700 DEDENQSPRSFQKTRHYFIAAVERLWDYGMSSPHVLNRAQSGSVQPKVQVQEBTD 1759
DB 1681 DEDENQSPRSFQKTRHYFIAAVERLWDYGMSSPHVLNRAQSGSVQPKVQVQEBTD 1740
QY 1760 GSFTQPLVKGELNEHGLGLGPVIRAEVDENIMVTFRNOASPRYSYSSLI SYEEHQROGA 1819
DB 1741 GSFTQPLVKGELNEHGLGLGPVIRAEVDENIMVTFRNOASPRYSYSSLI SYEEHQROGA 1800
QY 1820 EPRKNVFXNETKTYFWKVQHVMAPTDEDFCKAWAYTSDVDLEKDVHSGLIGPLLVCHT 1879
DB 1801 EPRKNVFXNETKTYFWKVQHVMAPTDEDFCKAWAYTSDVDLEKDVHSGLIGPLLVCHT 1860
QY 1880 NTLNPAHGRQVTVQBFALFTTIFDSTKSWYFTTENMERNCRAPCNQMEDPTFKENYRPHA 1939

DB 1861 NTLNPAHGRQVTVQBFALFTTIFDSTKSWYFTTENMERNCRAPCNQMEDPTFKENYRPHA 1920
QY 1940 INGYINDTLPGLVNAQDQRIRWYLLSMGNSNENIHSIHFSGHVFTYRKKEEYKVALYNLYP 1999
DB 1921 INGYINDTLPGLVNAQDQRIRWYLLSMGNSNENIHSIHFSGHVFTYRKKEEYKVALYNLYP 1980
QY 2000 GVFEVTEMLPSKAGIRWVECLIGEHLAGMSLFLVYNSKCOPTPLGMASGHRDFQITAS 2059
DB 1981 GVFEVTEMLPSKAGIRWVECLIGEHLAGMSLFLVYNSKCOPTPLGMASGHRDFQITAS 2040
QY 2060 GOYGOWAPKLAELHYSGSINAMSTKEPFSWIKVDLLAPMIHGIKTQGARQKPFSSLIYSQ 2119
DB 2041 GOYGOWAPKLAELHYSGSINAMSTKEPFSWIKVDLLAPMIHGIKTQGARQKPFSSLIYSQ 2100
QY 2120 FIIMYSLOGKKWQTVTRGNSTGFLVYFNGVDSGGIKHINFPPIIARIYRLHPTHSIRS 2179
DB 2101 FIIMYSLOGKKWQTVTRGNSTGFLVYFNGVDSGGIKHINFPPIIARIYRLHPTHSIRS 2160
QY 2180 TLRMELMGCDLNSCSPMLGMSKALSADAQITASSYFTNMFAWSPSKARLHLOQSNAR 2239
DB 2161 TLRMELMGCDLNSCSPMLGMSKALSADAQITASSYFTNMFAWSPSKARLHLOQSNAR 2220
QY 2240 PQVNNPKWLQVDFQTKMKTGVTVTQGVKSLLTSMYKFEFLISSQDGHQWTLFPQNGKV 2299
DB 2221 PQVNNPKWLQVDFQTKMKTGVTVTQGVKSLLTSMYKFEFLISSQDGHQWTLFPQNGKV 2280
QY 2300 KVFQGNQDSFTPVVNSLDPPLLTLYRLRHPOSVWHQIALRMEVLGCEAQDLY 2351
DB 2281 KVFQGNQDSFTPVVNSLDPPLLTLYRLRHPOSVWHQIALRMEVLGCEAQDLY 2332

RESULT 7

US-08-251-937A-4
; Sequence 4, Application US/08251937A
; Patent No. 5583209
; GENERAL INFORMATION:
; APPLICANT: Lollar, John S.
; TITLE OF INVENTION: Hybrid Human/Porcine Factor VIII
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Kilpatrick & Cody
; STREET: 1100 Peachtree Street
; CITY: Atlanta
; STATE: Georgia
; COUNTRY: US
; ZIP: 30309
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/251,937A
; FILING DATE: 31-MAY-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/864,004
; FILING DATE: 07-APR-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Pratt, John S.
; REGISTRATION NUMBER: 29,476
; REFERENCE/DOCKET NUMBER: EMU106DIV
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 404-815-6367
; TELEFAX: 404-815-6555
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2332 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear

MOLECULE TYPE: protein		95.0%; Score 12293; DB 1; Length 2332;	
HYPOTHETICAL: NO		Best Local Similarity 99.9%; Pred. No. 0;	
ANTI-SENSE: NO		Matches 2329; Conservative 1; Mismatches 2; Indels 0; Gaps 0;	
FRAGMENT TYPE: N-terminal			
ORIGINAL SOURCE:			
ORGANISM: Homo sapien			
TISSUE TYPE: Liver cDNA sequence			
US-08-251-937A-4			
QY	20	ATRRVYLGAVELSWYQMSDGLGELPVDARPPVPKSPFENTSVVYKTLFVDFTDHLN	79
DB	1	ATERVYLGAVELSWDQMSDGLGELPVDARPPVPKSPFENTSVVYKTLFVDFTDHLN	60
QY	80	IAXPRPMMGLLPTTQAEVYDVTWTLKXASHPVSLHAGVSYMKASGAGBYDQTSQ	139
DB	61	IAXPRPMMGLLPTTQAEVYDVTWTLKXASHPVSLHAGVSYMKASGAGBYDQTSQ	120
QY	140	RESEDDKVPFGSGHTYVQVILKNGPMASDPLCLTYSLSHVDLYKDLNSGLIGALLVCR	199
DB	121	REKEDKVPFGSGHTYVQVILKNGPMASDPLCLTYSLSHVDLYKDLNSGLIGALLVCR	180
QY	200	EGSLAKEKQTQLHKFILLFAVFDGKSWHSETKNSLMQDQDAASAPKWHVWNGVNR	259
DB	181	EGSLAKEKQTQLHKFILLFAVFDGKSWHSETKNSLMQDQDAASAPKWHVWNGVNR	240
QY	260	SPLGLICGCHRSYVHVIGMGTTPVHVSIFLEGHTFLVRNHRQASLEISPIELTAQTL	319
DB	241	SPLGLICGCHRSYVHVIGMGTTPVHVSIFLEGHTFLVRNHRQASLEISPIELTAQTL	300
QY	320	MDLQGLLCHISGHQDGNVAVKVDSCPEEPQLEKMNNEAEDYDDDLTDSMDVVRP	379
DB	301	MDLQGLLCHISGHQDGNVAVKVDSCPEEPQLEKMNNEAEDYDDDLTDSMDVVRP	360
QY	380	DDNSPFSFIQIRSVAKXGKPTWVHYIAAESEDDYAPLAPDDRYSYKSYLNGPQRIG	439
DB	361	DDNSPFSFIQIRSVAKXGKPTWVHYIAAESEDDYAPLAPDDRYSYKSYLNGPQRIG	420
QY	440	RYKKVRPMAYTETPKIRSAI CHESGILGPLYGEVGDITLLIFKNQASRPNIYPHGI	499
DB	421	RYKKVRPMAYTETPKIRSAI CHESGILGPLYGEVGDITLLIFKNQASRPNIYPHGI	480
QY	500	TDVRPLYSRRLPGVVKHLDKDFPILPGELPKYKWTIVVEDGPTKSDPCLTRYYSFVNME	559
DB	481	TDVRPLYSRRLPGVVKHLDKDFPILPGELPKYKWTIVVEDGPTKSDPCLTRYYSFVNME	540
QY	560	RLASGLIGPLLI CYKESVDQRGNQIMSDKRNVLFSVPDENRSWYLTENIQRFLENPAG	619
DB	541	RLASGLIGPLLI CYKESVDQRGNQIMSDKRNVLFSVPDENRSWYLTENIQRFLENPAG	600
QY	620	VQLEDPEFOASNMHSNGVVFDSQLSVCLHEVAYWYLLISGAQTDPLSVPSGTYFKH	679
DB	601	VQLEDPEFOASNMHSNGVVFDSQLSVCLHEVAYWYLLISGAQTDPLSVPSGTYFKH	660
QY	680	KMYVEDTLTLFPFSGETVFMSENPGIWLGCNSDFRNEGMALLKVSSCDKNTGDYE	739
DB	661	KMYVEDTLTLFPFSGETVFMSENPGIWLGCNSDFRNEGMALLKVSSCDKNTGDYE	720
QY	740	DSVEDISAYLLSKNNAIEPFSFONSHPSTRQKQFNATTIPENDIEKTDPFVFAHRTMP	799
DB	721	DSVEDISAYLLSKNNAIEPFSFONSHPSTRQKQFNATTIPENDIEKTDPFVFAHRTMP	780
QY	800	KIQNVSSDDLMLLRQSPTHGHSLSLDLQAKYETFPDDSPGAINSNLSSEWTHFRPQ	859
DB	781	KIQNVSSDDLMLLRQSPTHGHSLSLDLQAKYETFPDDSPGAINSNLSSEWTHFRPQ	840
QY	860	LHSGDMVFTPESGLQRLNEKLTGTTAATLKKLDFKVSSTSNLLSTIPSDNLAACTDN	919
DB	841	LHSGDMVFTPESGLQRLNEKLTGTTAATLKKLDFKVSSTSNLLSTIPSDNLAACTDN	900
QY	920	TSSLQPPSPMPVHYDSQDITLLFGKXSSPLTSGGPLSLSEENNKGKLLSGLMNSQESSW	979
DB	901	TSSLQPPSPMPVHYDSQDITLLFGKXSSPLTSGGPLSLSEENNKGKLLSGLMNSQESSW	960
QY	980	GKXVSTSGRLFKGKRAHPALLTKDNALFKVSISLLKTKNTKSNNSATNRKTHIDGSL	1039
DB	961	GKXVSTSGRLFKGKRAHPALLTKDNALFKVSISLLKTKNTKSNNSATNRKTHIDGSL	1020
QY	1040	LLENPSVWONTLESDBTEFKVTPLIHDMMLDKNATLRLNEMGNKTKTSSKMMVVOOK	1099
DB	1021	LLENPSVWONTLESDBTEFKVTPLIHDMMLDKNATLRLNEMGNKTKTSSKMMVVOOK	1080
QY	1100	XEGPDPDAQNMDGPFKMLFLPESARWIORHGNKSLNSGQCPKQIQLVSLGPKSVESG	1159
DB	1081	XEGPDPDAQNMDGPFKMLFLPESARWIORHGNKSLNSGQCPKQIQLVSLGPKSVESG	1140
QY	1160	QNFLESKNKVVYKGEFTKDVGLKEMVFPSSNRLFLTNLDNLHENNTHNQEKKEIEBK	1219
DB	1141	QNFLESKNKVVYKGEFTKDVGLKEMVFPSSNRLFLTNLDNLHENNTHNQEKKEIEBK	1200
QY	1220	KETLIQENNVLPQJHTVGTGKNFMKNLFLSTFRQNVGSGYAGYAPVLQDFPSLNDSTNR	1279
DB	1201	KETLIQENNVLPQJHTVGTGKNFMKNLFLSTFRQNVGSGYAGYAPVLQDFPSLNDSTNR	1260
QY	1280	TKKHTAHFSKGEENLEGLGNQTKQIYEKVACTTRISPTSQNFVQTSRKAALQKPL	1339
DB	1261	TKKHTAHFSKGEENLEGLGNQTKQIYEKVACTTRISPTSQNFVQTSRKAALQKPL	1320
QY	1340	PLEETELEKRIIIVDDTSTQWSKXMKHLTPSTTQIDYNEKEKGAITQSPSLCLTRSHI	1399
DB	1321	PLEETELEKRIIIVDDTSTQWSKXMKHLTPSTTQIDYNEKEKGAITQSPSLCLTRSHI	1380
QY	1400	POANSPLEIAKVSSPSPRIYTLFVLPQDNSSHLPAASYKKSQSGVSESHFLQAKK	1459
DB	1381	POANSPLEIAKVSSPSPRIYTLFVLPQDNSSHLPAASYKKSQSGVSESHFLQAKK	1440
QY	1460	NNLSLAILEMTGDQREYVSLGTSATNSVTYKVENTVLPKPDLPKTSKVELLPKVHI	1519
DB	1441	NNLSLAILEMTGDQREYVSLGTSATNSVTYKVENTVLPKPDLPKTSKVELLPKVHI	1500
QY	1520	YOKJLFPPTSNGSPCHLDLVEGSLLOGTGEGAIKWEANRPCKVPLRVATESSAKTPSK	1579
DB	1501	YOKJLFPPTSNGSPCHLDLVEGSLLOGTGEGAIKWEANRPCKVPLRVATESSAKTPSK	1560
QY	1580	LIDPLAWDNHYGTQIPKEBWKSOEKSPEKTA PKKDTILSLNACSNHAI AAINEGONXP	1639
DB	1561	LIDPLAWDNHYGTQIPKEBWKSOEKSPEKTA PKKDTILSLNACSNHAI AAINEGONXP	1620
QY	1640	EIEVTWAKOGRTERLCSQNPVLRKHOREITRTTILQSDQEEIDYDDTISVENMKEDFDY	1699
DB	1621	EIEVTWAKOGRTERLCSQNPVLRKHOREITRTTILQSDQEEIDYDDTISVENMKEDFDY	1680
QY	1700	DEDENOSPSFOKTHYFI AAVERLWDYGMSSPHVLNRACSGSVPOPKVVFQEFDD	1759
DB	1681	DEDENOSPSFOKTHYFI AAVERLWDYGMSSPHVLNRACSGSVPOPKVVFQEFDD	1740
QY	1760	GSFTQPLRGELNEHLGLGPYIRAEVEDNIMVTFRNQASRPYSFYSSLSIYEEDORQCA	1819
DB	1741	GSFTQPLRGELNEHLGLGPYIRAEVEDNIMVTFRNQASRPYSFYSSLSIYEEDORQCA	1800
QY	1820	EPKKNPVKNENKTYFWKVQHMAPTKDFDCKAWAYFSDVLEKDVHSGSLGLPLVCHT	1879
DB	1801	EPKKNPVKNENKTYFWKVQHMAPTKDFDCKAWAYFSDVLEKDVHSGSLGLPLVCHT	1860
QY	1880	NTLNPAGHQVTVQSEFALPFTTIFDET KSWYFTENMERNCRAPCNIQMEDPTKKNYRFFA	1939
DB	1861	NTLNPAGHQVTVQSEFALPFTTIFDET KSWYFTENMERNCRAPCNIQMEDPTKKNYRFFA	1920
QY	1940	INGYIMDTLPLGLVMAQDQIRWYLLSMGNSNENHSHIFSGHVFTVRKKEEYKWAYNLYP	1999
DB	1921	INGYIMDTLPLGLVMAQDQIRWYLLSMGNSNENHSHIFSGHVFTVRKKEEYKWAYNLYP	1980
QY	2000	GVFETVEMLPKSKAGIWRVBECLIGEHLHAGMSTLFLVYSNKCQTPGLGMASGHIRDPOITAS	2059

1981 CVFTEVEMLPKAGIWRVECEGLHAGMSTFLVYSNKCQOTPLGMASGHIROFQITAS 2040
2060 QYQGWAKIARLHYSGSINAWSTKEPESWIKVVDLLAPMIHIGIKTOGAROKFESSLYISQ 2119
2041 QYQGWAKIARLHYSGSINAWSTKEPESWIKVVDLLAPMIHIGIKTOGAROKFESSLYISQ 2100
2120 FIIMYSLGKKWQYRGNSTGFLMVFPGNVDSGGIKHINFPPIIARIRLHPHYSIRS 2179
2101 FIIMYSLGKKWQYRGNSTGFLMVFPGNVDSGGIKHINFPPIIARIRLHPHYSIRS 2160
2180 TLRMELMGCDLNSCWPLMGESKALSDAQITASSYFTNNFATWSPSKARLHLQGRSNAMR 2239
2161 TLRMELMGCDLNSCWPLMGESKALSDAQITASSYFTNNFATWSPSKARLHLQGRSNAMR 2220
2240 POWNPKEWLOVDFOKTKWKTGVITTOGVKSLTSMYKKEFLISSQDGHOWTLFPQNGKV 2299
2221 POWNPKEWLOVDFOKTKWKTGVITTOGVKSLTSMYKKEFLISSQDGHOWTLFPQNGKV 2280
2300 KVFQGNQDSFTPVNASLDPPLLTRYLRIHFQSWVHQAIRLMEVLGCEAQDLY 2351
2281 KVFQGNQDSFTPVNASLDPPLLTRYLRIHFQSWVHQAIRLMEVLGCEAQDLY 2332

RESULT 8
US-08-212-133A-2
; Sequence 2, Application US/08212133A
; Patent No. 5663060
; GENERAL INFORMATION:
; APPLICANT: Iollar, John S.
; APPLICANT: Runge, Marshall S.
; TITLE OF INVENTION: Hybrid Human/Animal Factor VIII
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Kilpatrick & Cody
; STREET: 100 Peachtree Street
; CITY: Atlanta
; STATE: Georgia
; COUNTRY: US
; ZIP: 30303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/212,133A
; FILING DATE: March 11, 1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/864,004
; FILING DATE: 07-APR-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Pabst, Patrea L.
; REGISTRATION NUMBER: 31,284
; REFERENCE/DOCKET NUMBER: EMU/76677
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 404-572-6508
; TELEFAX: 404-572-6555
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2332 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHEICAL: YES
; ANTI-SENSE: NO
; FRAGMENT TYPE: N-terminal
; ORIGINAL SOURCE:
; ORGANISM: Homo sapien
; TISSUE TYPE: Liver cDNA sequence

Query Match
Best Local Similarity 99.0%; Score 12293; DB 1; Length 2332;
Matches 2329; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
QY 20 ATRYYLGAVELSDWYQSDIGELPVDARPPRPVKFPFNTSVVYKTLFVEETDHLFN 79
DB 1 ATRYYLGAVELSDWYQSDIGELPVDARPPRPVKFPFNTSVVYKTLFVEETDHLFN 60
QY 80 IAKRPPMGLLGTIOAEVYDVTWITLKNMASHPVSLHVGVSYWKASGEAEVDDQTSQ 139
DB 61 IAKRPPMGLLGTIOAEVYDVTWITLKNMASHPVSLHVGVSYWKASGEAEVDDQTSQ 120
QY 140 REKEDDKVFPQGSHTYVQVLEKNGPMASDPLCLTYSLSHVDLVKDLNSGLIGALLVCR 199
DB 121 REKEDDKVFPQGSHTYVQVLEKNGPMASDPLCLTYSLSHVDLVKDLNSGLIGALLVCR 180
QY 200 EGSLAKEKTQTLHKPILLFAVFDGKSWHSETKNSLMQDRDAASARAWPKGHTVNGYVNR 259
DB 181 EGSLAKEKTQTLHKPILLFAVFDGKSWHSETKNSLMQDRDAASARAWPKGHTVNGYVNR 240
QY 260 SLPLGIGCHRSKSVYWHVIGMGTTPPEVHSIFLEGHTFLVRNHRQASLEISPTIFLTAQTL 319
DB 241 SLPLGIGCHRSKSVYWHVIGMGTTPPEVHSIFLEGHTFLVRNHRQASLEISPTIFLTAQTL 300
QY 320 MDLGOFLLSCHISSHOHDGMEAYVKVDSCPEEPOLRMKNNEEAEDYDDDLTDSMDVVRP 379
DB 301 MDLGOFLLSCHISSHOHDGMEAYVKVDSCPEEPOLRMKNNEEAEDYDDDLTDSMDVVRP 360
QY 380 DDNSPSFQIRSVAKKHPKTWVHYIAAEEDWDYAPLVLAPDDRYSYKSYLNNGPQIG 439
DB 361 DDNSPSFQIRSVAKKHPKTWVHYIAAEEDWDYAPLVLAPDDRYSYKSYLNNGPQIG 420
QY 440 RYKVKRPMAYTDETFKTRBAIOHESGILGPLLGEVGDITLLIFKQASPYMYPHGI 499
DB 421 RYKVKRPMAYTDETFKTRBAIOHESGILGPLLGEVGDITLLIFKQASPYMYPHGI 480
QY 500 TDVRPLYSRRLPKGVXGLKDPFLLPGBIFKWTVTVEDGPTKSDPPRCLITYYSFVNME 559
DB 481 TDVRPLYSRRLPKGVXGLKDPFLLPGBIFKWTVTVEDGPTKSDPPRCLITYYSFVNME 540
QY 560 RDLASGLIGPLLI CYKESVDQGNQIMSDKRNVLFSVFDENRSMWLTENTQRLPNPAG 619
DB 541 RDLASGLIGPLLI CYKESVDQGNQIMSDKRNVLFSVFDENRSMWLTENTQRLPNPAG 600
QY 620 VOLEDPEFOASNIMHSINGYVFDLSQLSVCLHEVAYWYILSIGATDPLSFVSGYTFKH 679
DB 601 VOLEDPEFOASNIMHSINGYVFDLSQLSVCLHEVAYWYILSIGATDPLSFVSGYTFKH 660
QY 680 KMVYEDTTLTLPFSGETVFMSEMPGLWILGCHNSDFRNRMGTALLKVSSCDKNTGDYIE 739
DB 661 KMVYEDTTLTLPFSGETVFMSEMPGLWILGCHNSDFRNRMGTALLKVSSCDKNTGDYIE 720
QY 740 DSYEDISAYLLSKNAIHPRSFSONSRHPSRQOFNATTIPENDIEKTDFFWFAHRTMP 799
DB 721 DSYEDISAYLLSKNAIHPRSFSONSRHPSRQOFNATTIPENDIEKTDFFWFAHRTMP 780
QY 800 KIONVSSSDIIMLLRQSPTPHGLSLSDLOEAKYETFDPSGALDSNNLSSEMTFRPQ 859
DB 781 KIONVSSSDIIMLLRQSPTPHGLSLSDLOEAKYETFDPSGALDSNNLSSEMTFRPQ 840
QY 860 LHSQDMVFTPEESGLQRLNEKLGTTAATELKLDFKVSSTSNLNLSTIPSDNLAAGTDN 919
DB 841 LHSQDMVFTPEESGLQRLNEKLGTTAATELKLDFKVSSTSNLNLSTIPSDNLAAGTDN 900
QY 920 TSSLGPPSPVHYDSDLTTFGRKSSPLTSGGGLSLSENNDSKLLIESGLMNSQBSW 979
DB 901 TSSLGPPSPVHYDSDLTTFGRKSSPLTSGGGLSLSENNDSKLLIESGLMNSQBSW 960
QY 980 GRNVSSSTESGELPKGRAHGPALLTKONALPKVSI LLLKTKNTKNSATNKRTHIDGSL 1039
DB 961 GRNVSSSTESGELPKGRAHGPALLTKONALPKVSI LLLKTKNTKNSATNKRTHIDGSL 1020

QY 1040 LIENSPVQWQVQKLESDTEFKVTPILHDMMLDKNATLRLNHNKNTTSSKNMENVQOK 1099
DB 1021 LIENSPVQWQVQKLESDTEFKVTPILHDMMLDKNATLRLNHNKNTTSSKNMENVQOK 1080
QY 1100 KEGPIPPDAQNDPMFFPKMLFLPESARWQTHGKNSLNSGGPSPKQLVSLGPKSVBEG 1159
DB 1081 KEGPIPPDAQNDPMFFPKMLFLPESARWQTHGKNSLNSGGPSPKQLVSLGPKSVBEG 1140
QY 1160 QNFLSEKQKVVVGKEFTKDVGLKMWPPSSNLFNLNLHENTHNOEKKI0EEIEK 1219
DB 1141 QNFLSEKQKVVVGKEFTKDVGLKMWPPSSNLFNLNLHENTHNOEKKI0EEIEK 1200
QY 1220 KETLIQENVLPQIHTVTGTFKFMKMLFLLSTRQNVESYDQAYAPVLDQFSLNDSNMR 1279
DB 1201 KETLIQENVLPQIHTVTGTFKFMKMLFLLSTRQNVESYDQAYAPVLDQFSLNDSNMR 1260
QY 1280 TKKHTAHSKKGEBENLEGLGNQTKQIVKXACTTRISNTSOQNFVQORSKALQKPL 1339
DB 1261 TKKHTAHSKKGEBENLEGLGNQTKQIVKXACTTRISNTSOQNFVQORSKALQKPL 1320
QY 1340 PLEETELEXRIIVDDTSTOWSNXMKHLTPSTLTQIDYNEKEKGAITQSPSLDCLTRSHSI 1399
DB 1321 PLEETELEXRIIVDDTSTOWSNXMKHLTPSTLTQIDYNEKEKGAITQSPSLDCLTRSHSI 1380
QY 1400 POANRSPPIAKVSPSPSTIRPIYLRVLQDNSSHLPAASYPKKDSGVQESSHFLQGAKK 1459
DB 1381 POANRSPPIAKVSPSPSTIRPIYLRVLQDNSSHLPAASYPKKDSGVQESSHFLQGAKK 1440
QY 1460 NNLISAILTLEMTGDOREVGSLGTATNSVTYKVENVLPKPDLPKTSKGVLELPKVHI 1519
DB 1441 NNLISAILTLEMTGDOREVGSLGTATNSVTYKVENVLPKPDLPKTSKGVLELPKVHI 1500
QY 1520 YOKDLFPTETSGRPHLDLVBGSLIQGTGEGAIKWNEANRPKVPFLRVATESSAKTPSK 1579
DB 1501 YOKDLFPTETSGRPHLDLVBGSLIQGTGEGAIKWNEANRPKVPFLRVATESSAKTPSK 1560
QY 1580 LLDPLAWDNHYGTQIPKESWKSQBSPEKTAFKKDTILSLNACSNHAIAINRGONKP 1639
DB 1561 LLDPLAWDNHYGTQIPKESWKSQBSPEKTAFKKDTILSLNACSNHAIAINRGONKP 1620
QY 1640 ELEVWAKOGRTERLCSQNPVVKHOREITRTTLOSQD0EIDYDDTISVENMKKEDFDIY 1699
DB 1621 ELEVWAKOGRTERLCSQNPVVKHOREITRTTLOSQD0EIDYDDTISVENMKKEDFDIY 1680
QY 1700 DEDENQSPASFQKTRHYPIAAVERLWDYGMSSSPFLNRNRAQSGSVPOFKKVVQEPFD 1759
DB 1681 DEDENQSPASFQKTRHYPIAAVERLWDYGMSSSPFLNRNRAQSGSVPOFKKVVQEPFD 1740
QY 1760 GSFTOPLVYRGEINEHLLGLGPVIRAEVDNIMVTFRNQASRPYSFYSSLSIYEEQORQA 1819
DB 1741 GSFTOPLVYRGEINEHLLGLGPVIRAEVDNIMVTFRNQASRPYSFYSSLSIYEEQORQA 1800
QY 1820 EPRKPFVKNETKTYFWKQVHNAWPKDFDCAMAYFSDVLEKDVHSGLIGPLLVCHT 1879
DB 1801 EPRKPFVKNETKTYFWKQVHNAWPKDFDCAMAYFSDVLEKDVHSGLIGPLLVCHT 1860
QY 1880 NTLNPAHQVTVQEFALPFTTIDETKSWYFTENMERNCRAPONTQMEDPTPKENYRPHA 1939
DB 1861 NTLNPAHQVTVQEFALPFTTIDETKSWYFTENMERNCRAPONTQMEDPTPKENYRPHA 1920
QY 1940 INGYINDTLPGVNAQDQIRWVLLSMGNGENILHSIHFSGHVFTVTKKKEYKVALYNLYP 1999
DB 1921 INGYINDTLPGVNAQDQIRWVLLSMGNGENILHSIHFSGHVFTVTKKKEYKVALYNLYP 1980
QY 2000 GVFETVEMLPFSKAGIWRVECLIGEHLHAGMSTLFLVYSNKCOTPLGWSGHIRDFOITAS 2059
DB 1981 GVFETVEMLPFSKAGIWRVECLIGEHLHAGMSTLFLVYSNKCOTPLGWSGHIRDFOITAS 2040
QY 2060 GYGQWAPKARLHYSGSINAWSTKBPFSWKVLLAPMIHGIKTQGARQRFSSLYISQ 2119
DB 2041 GYGQWAPKARLHYSGSINAWSTKBPFSWKVLLAPMIHGIKTQGARQRFSSLYISQ 2100
QY 2120 FIIVYSIDGKKWQYRGNSGTGTLMVFFGQVDSGSGIKENIFNPPIIARYIRLHPHYSIRS 2179

RESULT 9

US-08-474-503-2
; Sequence 2, Application US/08474503
; Patent No. 574446
; GENERAL INFORMATION:
; APPLICANT: Emory University
; TITLE OF INVENTION: Hybrid Human/Animal Factor VIII
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Kilpatrick & Cody
; STREET: 1100 Peachtree Street, Suite 2800
; CITY: Atlanta
; STATE: Georgia
; COUNTRY: US
; ZIP: 30309
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/474,503
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Pratt, John S.
; REGISTRATION NUMBER: 29,476
; REFERENCE/DOCKET NUMBER: EMO106CIP(3)
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 404-815-6500
; TELEFAX: 404-815-6555
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2332 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHEICAL: YES
; ANTI-SENSE: NO
; FRAGMENT TYPE: N-terminal
; ORIGINAL SOURCE:
; ORGANISM: Homo sapien
; TISSUE TYPE: Liver cDNA sequence
; US-08-474-503-2

Query Match 99.0%; Score 12293; DB 1; Length 2332;
Best Local Similarity 99.9%; Pred. NO. 0;
Matches 2329; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 20 ATERYVLGAVELSDVMQSDLGELPVDARFPPRPVPSPPNTSVYKTLFVEFTDHLFN 79
DB 1 ATERYVLGAVELSDVMQSDLGELPVDARFPPRPVPSPPNTSVYKTLFVEFTDHLFN 60
QY 80 IAKPRPPWMLGPTIQAEVDTVITLKNVASHVPSLHAGVSYWKASEGAEYDDQTSQ 139
DB 61 IAKPRPPWMLGPTIQAEVDTVITLKNVASHVPSLHAGVSYWKASEGAEYDDQTSQ 120

QY 140 REKEDDKVPPGGSHYVWQVILKENGPMASDPCLITYSYLSHVDLVKDLNSGLIGALLVCR 199
DB 121 REKEDDKVPPGGSHYVWQVILKENGPMASDPCLITYSYLSHVDLVKDLNSGLIGALLVCR 180
QY 200 EGS LAKEKTQTLHKFILLFAVDEGKSWSHSETKNSLMQDRDAASARAWPMHETVNGYVR 259
DB 181 EGS LAKEKTQTLHKFILLFAVDEGKSWSHSETKNSLMQDRDAASARAWPMHETVNGYVR 240
QY 260 SLPLGLIGCHRKSVYHVGWGTTPVHSIFLGGHTPLVRNHRQASLEISPIFLTAQTL 319
DB 241 SLPLGLIGCHRKSVYHVGWGTTPVHSIFLGGHTPLVRNHRQASLEISPIFLTAQTL 300
QY 320 YDLGQPLFLCHSHSHQHGDWEAYKVDSCPEBPQLRMKNEEAEDYDDLTDSEMDVVR 379
DB 301 MDLQGPFLFLCHSHSHQHGDWEAYKVDSCPEBPQLRMKNEEAEDYDDLTDSEMDVVR 360
QY 380 DDNDSPFTQIRSVAKKHPTWVHYIAAEEEDWDYAPLVLPADDSSYKQYLNNGPQRIG 439
DB 361 DDNDSPFTQIRSVAKKHPTWVHYIAAEEEDWDYAPLVLPADDSSYKQYLNNGPQRIG 420
QY 440 RYKXKVRFMAYUDEPFTKREAIQHSIGLGPLLYGRVGDTLIIIPKNOASRPYNIYBHI 499
DB 421 RYKXKVRFMAYUDEPFTKREAIQHSIGLGPLLYGEVGDTLIIIPKNOASRPYNIYBHI 480
QY 500 TVRPLYSRRLPKGVKHLKDPPLPGBEIPKYKWTVTVEDGPTKSDPCLTRYSSFVNME 559
DB 481 TVRPLYSRRLPKGVKHLKDPPLPGBEIPKYKWTVTVEDGPTKSDPCLTRYSSFVNME 540
QY 560 RDLASGLIGPLII CYKESVDQGNQIMSDKRVNILLFSVFDENRNYLNIORFLENPAG 619
DB 541 RDLASGLIGPLII CYKESVDQGNQIMSDKRVNILLFSVFDENRNYLNIORFLENPAG 600
QY 620 VQLEDEPEQASNMHSINGYVFDLSQLSVCLHEVAYWILSICAOTDFLSVFFSGYTPKH 679
DB 601 VQLEDEPEQASNMHSINGYVFDLSQLSVCLHEVAYWILSICAOTDFLSVFFSGYTPKH 660
QY 680 KMYVYEDTTLTLPFSGETVMSMENPCLWILGCHNSDFRNRGMTALLKVSSCDKNTGDYVE 739
DB 661 KMYVYEDTTLTLPFSGETVMSMENPCLWILGCHNSDFRNRGMTALLKVSSCDKNTGDYVE 720
QY 740 DSYEDISAYLLSKNNAI BRPSQNSRHPSTKQKQFNATTIPENDIEKTDPMFAHRTMP 799
DB 721 DSYEDISAYLLSKNNAI BRPSQNSRHPSTKQKQFNATTIPENDIEKTDPMFAHRTMP 780
QY 800 KIQNVSSDMLLRQSPFHGLSLSDLOEAKYETPSDDPSGALDSNNLSSEMTFRPQ 859
DB 781 KIQNVSSDMLLRQSPFHGLSLSDLOEAKYETPSDDPSGALDSNNLSSEMTFRPQ 840
QY 860 LHSQDMVFTPESGLOLRINEKLGTTAATLKLDPKVSSTNNLI STIPSNLAAGTDN 919
DB 841 LHSQDMVFTPESGLOLRINEKLGTTAATLKLDPKVSSTNNLI STIPSNLAAGTDN 900
QY 920 TSSLGPPNPNVHYDSQDPTTLFGKSSPLTESGGLPSISENNDSKLLSGLMNSQESSW 979
DB 901 TSSLGPPNPNVHYDSQDPTTLFGKSSPLTESGGLPSISENNDSKLLSGLMNSQESSW 960
QY 980 GKNVSSSTESGRIFKGRAGPALLTKDNALFKVSI SLKTKNTKNSATNRKTHIDGPSL 1039
DB 961 GKNVSSSTESGRIFKGRAGPALLTKDNALFKVSI SLKTKNTKNSATNRKTHIDGPSL 1020
QY 1040 LIENSFSPWQNLIESDTEPKKTPDLIHRMLMDKXATALRLNHNKNTTSSKQMEVVOOK 1099
DB 1021 LIENSFSPWQNLIESDTEPKKTPDLIHRMLMDKXATALRLNHNKNTTSSKQMEVVOOK 1080
QY 1100 KEGP-PPDAQNDMSFFKVLRLPESARWIORTHGKNSLNSGGPSPKQIVSLGPESKVEG 1159
DB 1081 KEGP-PPDAQNDMSFFKVLRLPESARWIORTHGKNSLNSGGPSPKQIVSLGPESKVEG 1140
QY 1160 QNFLEKKNVVVGKEFTKDVGLKEMVFPSSRNLFITLNDLHNHNTNHOEKKIOBEIEK 1219
DB 1141 QNFLEKKNVVVGKEFTKDVGLKEMVFPSSRNLFITLNDLHNHNTNHOEKKIOBEIEK 1200

QY 1220 KETLIQENWVLPJQHTVTGKNFMKNLFLSTRQNVESYDGAYAPVLODFRSLNDSTNR 1279
DB 1201 KETLIQENWVLPJQHTVTGKNFMKNLFLSTRQNVESYDGAYAPVLODFRSLNDSTNR 1260
QY 1280 TKKHTAHFSKXGEENLEGLNQTKQIIVEKYACTTTRISPTNSQQNFVTRQSKRALKQFRL 1339
DB 1261 TKKHTAHFSKXGEENLEGLNQTKQIIVEKYACTTTRISPTNSQQNFVTRQSKRALKQFRL 1320
QY 1340 ZLEBEVELEKRIIVDDTSTQWSKNMKHLTJESLITQIDYNEKEKGAITQSPSLDCLTRSHI 1399
DB 1321 ZLEBEVELEKRIIVDDTSTQWSKNMKHLTJESLITQIDYNEKEKGAITQSPSLDCLTRSHI 1380
QY 1400 POANESPIPIAKVSSFPSPRIPIYLRVLFQDNSSHLPAASVRKDKSGVQESSHFLOGAKK 1459
DB 1381 POANESPIPIAKVSSFPSPRIPIYLRVLFQDNSSHLPAASVRKDKSGVQESSHFLOGAKK 1440
QY 1460 NNLSLAITLLEMTGQREVGSIGTSATNSVTVYKVENTVLPKPDLPKTSKGYELLPKVHI 1519
DB 1441 NNLSLAITLLEMTGQREVGSIGTSATNSVTVYKVENTVLPKPDLPKTSKGYELLPKVHI 1500
QY 1520 YOKDLPFTTSTNGSPGHLDLVEGSLLOCTEGAIKWNBANRPKVPLVATBESSAKTFSK 1579
DB 1501 YOKDLPFTTSTNGSPGHLDLVEGSLLOCTEGAIKWNBANRPKVPLVATBESSAKTFSK 1560
QY 1580 LLDPLAWDNHXYGTQIIPKEEWSQESPEKTAPKKDITILSNACSNHAI AAINEGONKP 1639
DB 1561 LLDPLAWDNHXYGTQIIPKEEWSQESPEKTAPKKDITILSNACSNHAI AAINEGONKP 1620
QY 1640 ELEVTAQGRTEELCSQNPVVLKHQBIEITRTTLOSQEBIDYDDTISVENMKEDFDIY 1699
DB 1621 ELEVTAQGRTEELCSQNPVVLKHQBIEITRTTLOSQEBIDYDDTISVENMKEDFDIY 1680
QY 1700 DEDENCSRSPQKTRHYFIIAAVEBLWDYGMSSSPHVLNRNAQSGSVPOFKKVVFOETD 1759
DB 1681 DEDENCSRSPQKTRHYFIIAAVEBLWDYGMSSSPHVLNRNAQSGSVPOFKKVVFOETD 1740
QY 1760 GSFTPLRYGELNEHGLGPGYIRAEVEDNTMVTRNOASRPSPSYSSLSIYEEEDOROGA 1819
DB 1741 GSFTPLRYGELNEHGLGPGYIRAEVEDNTMVTRNOASRPSPSYSSLSIYEEEDOROGA 1800
QY 1820 EPRKNFVKPNETKIYFKVQVQHMAPTDEFPDCKAWAYFSDVDLEKDHSGLIGPLLVCHT 1879
DB 1801 EPRKNFVKPNETKIYFKVQVQHMAPTDEFPDCKAWAYFSDVDLEKDHSGLIGPLLVCHT 1860
QY 1880 NTLNPAHGRQVTVCEPFFIPIETKSWYETENNERNCRAPCNQIOMEDPTFKENYRPHA 1939
DB 1861 NTLNPAHGRQVTVCEPFFIPIETKSWYETENNERNCRAPCNQIOMEDPTFKENYRPHA 1920
QY 1940 INGYIMDTPLGLVNAQDQIRWYLLSMGSNENIHSIHFSGHVFTVRKKEEYKMALYNLYP 1999
DB 1921 INGYIMDTPLGLVNAQDQIRWYLLSMGSNENIHSIHFSGHVFTVRKKEEYKMALYNLYP 1980
QY 2000 GVFEIVEMLPKAGIWRVCELIJGELHAGMSTLFLVSNKQOTPLGMASGHIRPOITAS 2059
DB 1981 GVFEIVEMLPKAGIWRVCELIJGELHAGMSTLFLVSNKQOTPLGMASGHIRPOITAS 2040
QY 2060 GOYGOWAPKLARLYVSGSINAWSTKEPSPWIKVLLAPMI IHGKTQCAROKFSSLYISQ 2119
DB 2041 GOYGOWAPKLARLYVSGSINAWSTKEPSPWIKVLLAPMI IHGKTQCAROKFSSLYISQ 2100
QY 2120 FIIMYSLDGKKWQYTRGNSTGTLVFFGNVDSSGIKHNIPNPPIIARYIRLEPHYSIRS 2179
DB 2101 FIIMYSLDGKKWQYTRGNSTGTLVFFGNVDSSGIKHNIPNPPIIARYIRLEPHYSIRS 2160
QY 2180 TLRMELMGCDLNSCMLPGMESKASIDAOITASSYFTNMWFTWSPSKARLHLOQRSHAWR 2239
DB 2161 TLRMELMGCDLNSCMLPGMESKASIDAOITASSYFTNMWFTWSPSKARLHLOQRSHAWR 2220
QY 2240 PQVNNPKEWLOVDQKTMKVTVTTQGVKSLTSMYKVEFLISSQDGHQWTLFFQNGKV 2299
DB 2221 PQVNNPKEWLOVDQKTMKVTVTTQGVKSLTSMYKVEFLISSQDGHQWTLFFQNGKV 2280
QY 2300 KVFOGQDSFTFVNVNSLDPPLLTTRYLRHPOSWHQIALRMEVLGCEBAQDLY 2351

Db 2261 KVFQGNDSPTVNSLDPPLLRYLRHPQSWHQIALRNEVLGCEAQDLY 2332

RESULT 10

US-08-670-707A-2
; Sequence 2, Application US/08670707A
; Patent No. 5859204
; GENERAL INFORMATION:
; APPLICANT: Lollar, John S.
; TITLE OF INVENTION: Hybrid Human/Animal Factor VIII
; NUMBER OF SEQUENCES: 40
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Greenlee, Winner and Sullivan, P.C.
; STREET: 5370 Manhattan Circle Suite 201
; CITY: Boulder
; STATE: Colorado
; COUNTRY: USA
; ZIP: 80303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/670,707A
; FILING DATE: 26-JUN-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: WO PCT/US94/13200
; FILING DATE: 15-NOV-1994

; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/212,133
; FILING DATE: 11-MAR-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/864,004
; FILING DATE: 07-APR-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Greenlee, Lorraine L.
; REGISTRATION NUMBER: 27,894
; REFERENCE/DOCKET NUMBER: 75-95F
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 303/499-8080
; TELEFAX: 303/499-8089

; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2332 amino acids
; TYPE: amino acid

; STRANDEDNESS: single
; TOPOLOGY: not relevant
; MOLECULE TYPE: protein
; HYPOTHETICAL: YES
; ANTI-SENSE: NO
; FRAGMENT TYPE: N-terminal
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; TISSUE TYPE: Liver

US-08-670-707A-2

Query Match 99.0%; Score 12293; DB 2; Length 2332;
Best Local Similarity 99.9%; Prod. No. 0;
Matches 2329; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 20 ATRRYILGAVELSDWYNQSDLGELPVDARPPVPKSPFPNTSVVYKTLFVFTDHLFN 79

Db 1 ATRRYILGAVELSDWYNQSDLGELPVDARPPVPKSPFPNTSVVYKTLFVFTDHLFN 60

QY 80 IAKRPWMLGLGTTQAEYVDVTWITLKNWASHPVSLHAGVSYWYKASGAEYDDQTSQ 139

Db 61 IAKRPWMLGLGTTQAEYVDVTWITLKNWASHPVSLHAGVSYWYKASGAEYDDQTSQ 120

QY 140 REKEDDKVFPGGSHTYVWQVLKENGPMASDPLCLITYSLSHVDLVKDLNSGLIGALLVCR 199

Db 121 REKEDDKVFPGGSHTYVWQVLKENGPMASDPLCLITYSLSHVDLVKDLNSGLIGALLVCR 180

QY 200 EGS LAKEKTQTLLKZILLFAVFEDEGKSWHSEFKNSILMORDDAASARAWPKMTVNGYVNR 259

Db 181 EGS LAKEKTQTLLKZILLFAVFEDEGKSWHSEFKNSILMORDDAASARAWPKMTVNGYVNR 240

QY 260 SLPLGLIGCHRKSVYWHVIGWGTTPEVHSIFLECHTFLVRNHRQASLEISPIFLTAQTLL 319

Db 241 SLPLGLIGCHRKSVYWHVIGWGTTPEVHSIFLECHTFLVRNHRQASLEISPIFLTAQTLL 300

QY 320 MDLGQFLLSCHISSHQHGWAEYKVDSCPEEPQIRMKNNEAEYDDDLTSEMDDVVRP 379

Db 301 MDLGQFLLSCHISSHQHGWAEYKVDSCPEEPQIRMKNNEAEYDDDLTSEMDDVVRP 360

QY 380 DDNDSPTQIBSVAKKPKTWVHYIAABEEDWDYAPLVLAPDDRYSKYKSOYLNNGPQRTG 439

Db 361 DDNDSPTQIBSVAKKPKTWVHYIAABEEDWDYAPLVLAPDDRYSKYKSOYLNNGPQRTG 420

QY 440 RYKVKVRFMAYTDETFKTRERACHESGILGPLLYGEVGTLLIIFKNQASRPYNIYPHGT 499

Db 421 RYKVKVRFMAYTDETFKTRERACHESGILGPLLYGEVGTLLIIFKNQASRPYNIYPHGT 480

QY 500 TDVRPLYSRRLPKGVKHLKDFILPGEIPIKYKWTVTVEDGPTKSDPRCLTRYSSVNMME 559

Db 481 TDVRPLYSRRLPKGVKHLKDFILPGEIPIKYKWTVTVEDGPTKSDPRCLTRYSSVNMME 540

QY 560 RDLASGLIGPLLI CYKESVDQRGNQIMSKRNVILFSVPDENRSWYLTENIQRFJNPAG 619

Db 541 RDLASGLIGPLLI CYKESVDQRGNQIMSKRNVILFSVPDENRSWYLTENIQRFJNPAG 600

QY 620 VQLEDEPQASNMHSINGYVFDLSQLSVCLHEVAYWYILSIGAQTDFLSPVPSGYTFKH 679

Db 601 VQLEDEPQASNMHSINGYVFDLSQLSVCLHEVAYWYILSIGAQTDFLSPVPSGYTFKH 660

QY 680 KQVYEDTLFLPESGETVMSMENPGLWILGCHNSDFRNRGTMALLKVSSCDKNTGDYIE 739

Db 661 KQVYEDTLFLPESGETVMSMENPGLWILGCHNSDFRNRGTMALLKVSSCDKNTGDYIE 720

QY 740 DSYEDISAYLLSKNNAIEPRSPQNSRHPSTKQKNATTIIPENDIEKTDPMFAHRTMP 799

Db 721 DSYEDISAYLLSKNNAIEPRSPQNSRHPSTKQKNATTIIPENDIEKTDPMFAHRTMP 780

QY 800 KIQNVSSSDLLMLRQSPPHGLSLSDLOEAKYETPSDDPSPCAIDSNNSLSMTHFRPQ 859

Db 781 KIQNVSSSDLLMLRQSPPHGLSLSDLOEAKYETPSDDPSPCAIDSNNSLSMTHFRPQ 840

QY 860 LHSQGMVFTPSGLQLRLNEKLGTTAATLKLDEKSVSTNNLLISTIPSDNLAAGTGN 919

Db 841 LHSQGMVFTPSGLQLRLNEKLGTTAATLKLDEKSVSTNNLLISTIPSDNLAAGTGN 900

QY 920 TSSLGPPSPMEVHYDSOLDTTLFGKXSSPLTESGGPLSLSEENNDKLLBSGLMNSQESSW 979

Db 901 TSSLGPPSPMEVHYDSOLDTTLFGKXSSPLTESGGPLSLSEENNDKLLBSGLMNSQESSW 960

QY 980 GKNVSTESGRILFKGKRAHGPALLTKDNALFKVSISSLTKNTKTSNNSANRKHIDGPSL 1039

Db 961 GKNVSTESGRILFKGKRAHGPALLTKDNALFKVSISSLTKNTKTSNNSANRKHIDGPSL 1020

QY 1040 LIENSPSVWONILESDETEPKVTPLIHDRMLDKNATLRLNHNMSKNTSSKNMNVQOK 1099

Db 1021 LIENSPSVWONILESDETEPKVTPLIHDRMLDKNATLRLNHNMSKNTSSKNMNVQOK 1080

QY 1100 KSGPPIPPDAQNDPMGFFKMLFLPESARWTQRTGKNSLNSGGSPKQLVSLGPBKSVEG 1159

Db 1081 KSGPPIPPDAQNDPMGFFKMLFLPESARWTQRTGKNSLNSGGSPKQLVSLGPBKSVEG 1140

QY 1160 QNFLSEKXKVVYKGBFTKDVGLKEMWFPSSNLFNLNLDNLHNHNTNQRKKEIEIEK 1219

Db 1141 QNFLSEKXKVVYKGBFTKDVGLKEMWFPSSNLFNLNLDNLHNHNTNQRKKEIEIEK 1200

QY 1220 KETLQENNVVLQIHTVTGTFKMFKNLFLSLTRQNVESYDGAAYAVLQDPRSLNDSTNR 1279

Db 1201 KETLQENNVVLQIHTVTGTFKMFKNLFLSLTRQNVESYDGAAYAVLQDPRSLNDSTNR 1260

1280 TKKHTAHFSKKEBENLEGLNQTQKQVEKYACTTRISNTSQONFVTPORSKALQKREL 1339
Db TKKHTAHFSKKEBENLEGLNQTQKQVEKYACTTRISNTSQONFVTPORSKALQKREL 1320
QY PLEETEELKRIIIVDTSTOWSKMKHLPSTLTQIDYNEKEKGAITQSPLSCLTRSHSI 1399
Db PLEETEELKRIIIVDTSTOWSKMKHLPSTLTQIDYNEKEKGAITQSPLSCLTRSHSI 1380
QY POANSSPIPIAKVSPSPSR2YLRVLPQDNSSHLPAASYKEDSGVQESSHFLQGAKK 1459
Db POANSSPIPIAKVSPSPSR2YLRVLPQDNSSHLPAASYKEDSGVQESSHFLQGAKK 1440
QY NNLSLAILTLEMTGTQREYVGLTSGATNSVTYKQVNTVLPKPDLPKSGKVELLPKVHI 1519
Db NNLSLAILTLEMTGTQREYVGLTSGATNSVTYKQVNTVLPKPDLPKSGKVELLPKVHI 1500
QY YOKDLPPTETNSGSHLDLVEGSLLOQTEGAIKWNEANRPKVPLRVATESSAKTFSK 1579
Db YOKDLPPTETNSGSHLDLVEGSLLOQTEGAIKWNEANRPKVPLRVATESSAKTFSK 1560
QY LLDPLAWNHGYTOIPKEWKSQKSPKTAPEKXDTILSNACSNHAIAINEGQKNP 1639
Db LLDPLAWNHGYTOIPKEWKSQKSPKTAPEKXDTILSNACSNHAIAINEGQKNP 1620
QY EIEVTWAKQGRTERLCSONPPVLPKHQREBITRTTLQSQDEEIDYDDTISVENMKKEDFDIY 1699
Db EIEVTWAKQGRTERLCSONPPVLPKHQREBITRTTLQSQDEEIDYDDTISVENMKKEDFDIY 1680
QY DEDENQSPRSQKTRHYFIAAVERLWYDGMSSSPHVLNRNAQSGSVPOFKVQVQEFDT 1759
Db DEDENQSPRSQKTRHYFIAAVERLWYDGMSSSPHVLNRNAQSGSVPOFKVQVQEFDT 1740
QY CSFTOPLYRGELNEHLGLLGPYIRAEVDENIMVTFRNOASRYSYSSLSIYEEBQROGA 1819
Db CSFTOPLYRGELNEHLGLLGPYIRAEVDENIMVTFRNOASRYSYSSLSIYEEBQROGA 1800
QY EPRKNFVKNETKTYFWKQVHNAPTKDEFDCKAWAYFSDVLEKDVHSGLIGPLLVCHT 1879
Db EPRKNFVKNETKTYFWKQVHNAPTKDEFDCKAWAYFSDVLEKDVHSGLIGPLLVCHT 1860
QY NTLNPAHGRQVTVQSFALFFITFDTKWYFTENMERNCRAPCNQMEDPTFKENYRHEA 1939
Db NTLNPAHGRQVTVQSFALFFITFDTKWYFTENMERNCRAPCNQMEDPTFKENYRHEA 1920
QY INGYIMDTLPLGLVMAQDQRIKRYLLSMGSNENIHSIHFSGHVFTVRKKEBYKMALYNLYP 1999
Db INGYIMDTLPLGLVMAQDQRIKRYLLSMGSNENIHSIHFSGHVFTVRKKEBYKMALYNLYP 1980
QY GVFTVEMLPSKAGTWRVECLTIGELHAGMSTLFLVYGNKQCTPLGMASGHIRDPQITAS 2059
Db GVFTVEMLPSKAGTWRVECLTIGELHAGMSTLFLVYGNKQCTPLGMASGHIRDPQITAS 2040
QY GOYQWAPKALRHLHYSGSNINAWSTKEPSWIKVDLAPMIHGIKTQGARQKFSLLYISQ 2119
Db GOYQWAPKALRHLHYSGSNINAWSTKEPSWIKVDLAPMIHGIKTQGARQKFSLLYISQ 2100
QY FIIMYSLDGKKWQYRGNSTGLTVFFGNVDSGGIKHNIENPPIIARVIRLHPHYSIRS 2179
Db FIIMYSLDGKKWQYRGNSTGLTVFFGNVDSGGIKHNIENPPIIARVIRLHPHYSIRS 2160
QY TIRMLMGCDLNSCMLPGMESKA-SDAQITASSYFTNMFATWSPSKARLHQGRSNAWR 2239
Db TIRMLMGCDLNSCMLPGMESKA-SDAQITASSYFTNMFATWSPSKARLHQGRSNAWR 2220
QY POVNNPKEWLQVDFQKTKWYGTGVTTCQVKSLTSMYVKEFLISSQDGEQWTLFQNGKV 2299
Db POVNNPKEWLQVDFQKTKWYGTGVTTCQVKSLTSMYVKEFLISSQDGEQWTLFQNGKV 2280
QY KVFQGNQDSFTPVVNSLDPPLLTIRLIRHPQSWHVOIALRMEVLGCEAQDLY 2351
Db KVFQGNQDSFTPVVNSLDPPLLTIRLIRHPQSWHVOIALRMEVLGCEAQDLY 2332

RESULT 11
JS-09-037-601-2
; Sequence 2, Application US/09037601
; Patent No. 6190371
; GENERAL INFORMATION:
; APPLICANT: Lollar, John S.
; TITLE OF INVENTION: Hybrid Human/Animal Factor VIII
; NUMBER OF SEQUENCES: 40
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Greenlee, Winner and Sullivan, P.C.
; STREET: 5370 Manhattan Circle Suite 201
; CITY: Boulder
; STATE: Colorado
; COUNTRY: USA
; ZIP: 80303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/037,601
; FILING DATE: 26-JUN-1996
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: WO PCT/US94/13200
; FILING DATE: 15-NOV-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/212,133
; FILING DATE: 11-MAR-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/864,004
; FILING DATE: 07-APR-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Ferber, Donna M.
; REGISTRATION NUMBER: 33,878
; REFERENCE/DOCKET NUMBER: 75-95F
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 303/499-8080
; TELEFAX: 303/499-8089
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2332 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: not relevant
; MOLECULE TYPE: protein
; HYPOTHEetical: YES
; ANTI-SENSE: NO
; FRAGMENT TYPE: N-terminal
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; TISSUE TYPE: Liver
; JS-09-037-601-2

Query Match 99.0%; Score 12293; DB 3; Length 2332;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 2329; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
QY 20 ATRRYLGAVELSDYMQSDLGELPVDARPPRPVKSPFNTSVVYKTLFVEFTDHLFN 79
Db 1 ATRRYLGAVELSDYMQSDLGELPVDARPPRPVKSPFNTSVVYKTLFVEFTDHLFN 60
QY 80 IAKRPPFMGLGPTIQAEVYDVTWITLKNMASHFVSLHVGVSYWKASGAEYDDQTSQ 139
Db 61 IAKRPPFMGLGPTIQAEVYDVTWITLKNMASHFVSLHVGVSYWKASGAEYDDQTSQ 120
QY 140 REKEDDKVFPGSGSHYVQVVKENGPMASDPLCTYSYLSHVDLVKDLNSGLIGALLVCR 199
Db 121 REKEDDKVFPGSGSHYVQVVKENGPMASDPLCTYSYLSHVDLVKDLNSGLIGALLVCR 180
QY 200 EGSIAKERTQTLHKFILLFAVFDGKSWHSETKNSLMQDRDASARAWPKMTVNGYVNR 259

181 EGS LAKEKQT LHKFILLFAVPEDEKSWSETKNSLMQDRDAASARAWPKWHTUNGYNR 240
260 SLPGILGCHRKSVYWHVIMGWTTPVHSIFLBGHTFLVRNHRQASLEISPIFLTAQTL 319
241 SLPGILGCHRKSVYWHVIMGWTTPVHSIFLBGHTFLVRNHRQASLEISPIFLTAQTL 300
320 MDLQGPLLSCHTSSHOHDGMEAYVKVDSPPEBQLRMKNNEBAEDYDDDLTISEMDVVF 379
301 MDLQGPLLRCHTSSHOHDGMEAYVKVDSPPEBQLRMKNNEBAEDYDDDLTISEMDVVF 360
380 DDNSPSFTQIRSVAKKHKTWWHYIAABEEDWDYAPLVADDDRSYKGOYLNNQQRIG 439
361 DDNSPSFTQIRSVAKKHKTWWHYIAABEEDWDYAPLVADDDRSYKGOYLNNQQRIG 420
440 RYKVKVRFMAYTDEFTKREAOHESGILGPLLYGEVGTLLIIFKQOASRPVNIYPHGI 499
421 RYKVKVRFMAYTDEFTKREAOHESGILGPLLYGEVGTLLIIFKQOASRPVNIYPHGI 480
500 TDVRPLYSRRLPKGVKHLXDFPILPGEIPKYKWTVTVEGPTKSDPRCLTRYYSFVWNE 559
481 TDVRPLYSRRLPKGVKHLXDFPILPGEIPKYKWTVTVEGPTKSDPRCLTRYYSFVWNE 540
560 RLASGLIPLIICYKESVDQRNQIMSDKRNWILFSVFDENRSWYLTENIQRFPLNPAG 619
541 RLASGLIPLIICYKESVDQRNQIMSDKRNWILFSVFDENRSWYLTENIQRFPLNPAG 600
620 VOLEDEPOASNIHMSINGVFDLSQLSVCLHEVAYWYILSICAOTDFLSVPSGVTYRKH 679
601 VOLEDEPOASNIHMSINGVFDLSQLSVCLHEVAYWYILSICAOTDFLSVPSGVTYRKH 660
680 XMYEDTLTLPFSGETVMSMENPGWLILGCHNSDFRNRMGTALLKVSSCDKNTGDYVE 739
661 XMYEDTLTLPFSGETVMSMENPGWLILGCHNSDFRNRMGTALLKVSSCDKNTGDYVE 720
740 DSYEDISAYLLSKNAIBRPSQNSRHSSTQKQFNATTIPENDIEKTDHPFAHRTMP 799
721 DSYEDISAYLLSKNAIBRPSQNSRHSSTQKQFNATTIPENDIEKTDHPFAHRTMP 780
800 KTVNSSSDLLMLLROSPTPHGLSLSDLOBAKYETFSDDPSGAISSNLSSEMTFRPQ 859
781 KTVNSSSDLLMLLROSPTPHGLSLSDLOBAKYETFSDDPSGAISSNLSSEMTFRPQ 840
860 LHSXGDMVTPSGIQLRLNEKLTGTAATLKLDPKVSSTNNLIISTIPSDNLAAGTEN 919
841 LHSXGDMVTPSGIQLRLNEKLTGTAATLKLDPKVSSTNNLIISTIPSDNLAAGTEN 900
920 TSSLGPPSPMPVHYDSOLDTTLFGKXSSPLTESGGPLSLSEENDDSKLLESGLMNOQESSW 979
901 TSSLGPPSPMPVHYDSOLDTTLFGKXSSPLTESGGPLSLSEENDDSKLLESGLMNOQESSW 960
980 GKNVSTESGRFLFKGKRAHGFALLTKDNALFKVISLSLLKTNKTSNNSATNRKTHIDGPSL 1039
961 GKNVSTESGRFLFKGKRAHGFALLTKDNALFKVISLSLLKTNKTSNNSATNRKTHIDGPSL 1020
1040 LIENSFVWQNLIESDTEPKVTPLIHDRMLMDKNATLRLNHNKNTSSKNMVMVQOK 1099
1021 LIENSFVWQNLIESDTEPKVTPLIHDRMLMDKNATLRLNHNKNTSSKNMVMVQOK 1080
1100 KEGPIPPDAQNPDMPFKMLFLPESARWIORTHGKNSLNSGQSPKQIVSIGPEKSVEG 1159
1081 KEGPIPPDAQNPDMPFKMLFLPESARWIORTHGKNSLNSGQSPKQIVSIGPEKSVEG 1140
1160 QNFLESEKNVYVKGGEFTKDVGLKEMVFPSSRNLFNLNLFENNTHOEBKIOBETEK 1219
1141 QNFLESEKNVYVKGGEFTKDVGLKEMVFPSSRNLFNLNLFENNTHOEBKIOBETEK 1200
1220 KETLIDENWVLPQIHVTGTQKFMKNLFLLSRQNVVEGSDYDGAAPVLQDFSLNDSTNR 1279
1201 KETLIDENWVLPQIHVTGTQKFMKNLFLLSRQNVVEGSDYDGAAPVLQDFSLNDSTNR 1260
1280 TKKHTAHFSKKEEENLEGLGNQTKQIVKQACTRISNTSOONFVTOXSRLAKQFRL 1339

RESULT 12

US-09-313-179-2

1261 TKKHTAHFSKKEEENLEGLGNQTKQIVKQACTRISNTSOONFVTOXSRLAKQFRL 1320
1340 P-EBTELKRIIVDDTSTQMSXMKHLT2STLTQTDYNEKEKGAITOSPDLCLTRSHSI 1339
1321 P-EBTELKRIIVDDTSTQMSXMKHLT2STLTQTDYNEKEKGAITOSPDLCLTRSHSI 1380
1400 POANRSPPIAKVSSFPPIRPIYLRVLFDQNSSHLPAA5YRKDKSGVQESSHFLQGAKK 1459
1381 POANRSPPIAKVSSFPPIRPIYLRVLFDQNSSHLPAA5YRKDKSGVQESSHFLQGAKK 1440
1460 NNL5LAITLLEMTGQREVGLSGT5ATNSVYKVKVENTVLPKDLPKTSKGVELLPKWHI 1519
1441 NNL5LAITLLEMTGQREVGLSGT5ATNSVYKVKVENTVLPKDLPKTSKGVELLPKWHI 1500
1520 YOKDIPPTETNSGSPGHLDLVEGSLLOGTGEGA1KWEANRPGKVPFLRVATESSAKTFSK 1579
1501 YOKDIPPTETNSGSPGHLDLVEGSLLOGTGEGA1KWEANRPGKVPFLRVATESSAKTFSK 1560
1580 LLDPLAWNHNHYGTQIPKEEWSQEKSPKTAPEKXDTILSLNACSNHAI1A1N5GQNK 1639
1561 LLDPLAWNHNHYGTQIPKEEWSQEKSPKTAPEKXDTILSLNACSNHAI1A1N5GQNK 1620
1640 E1EVTWAKQGRTEPLCSQNPVPLKHOREITRTTLTQSQOEEIDYD2TISVEMKKEDFIY 1699
1621 E1EVTWAKQGRTEPLCSQNPVPLKHOREITRTTLTQSQOEEIDYD2TISVEMKKEDFIY 1680
1700 DEDENOSPRSFOKTRHYFIAAVERLWDYGMSSSPHVLNRQAQSGVPQFKVVFCEFTD 1759
1681 DEDENOSPRSFOKTRHYFIAAVERLWDYGMSSSPHVLNRQAQSGVPQFKVVFCEFTD 1740
1760 GSFTQPLYRGELNEHLGLGPYIRAEVEDCNMTWTFPRNOASRPYSYSSLSYEEQROGA 1819
1741 GSFTQPLYRGELNEHLGLGPYIRAEVEDCNMTWTFPRNOASRPYSYSSLSYEEQROGA 1800
1820 EBRKRVKPNETKTYFWKQHHMADTKDBFPCKAWAYFSDVDLEKDVHSGLIGPLLVCHT 1879
1801 EBRKRVKPNETKTYFWKQHHMADTKDBFPCKAWAYFSDVDLEKDVHSGLIGPLLVCHT 1860
1880 NTLNPAHGRQVTVQ3FALLPFTIFDETKSWYFTENNERNCAPCNIQMEDPFPKNVRPHA 1939
1861 NTLNPAHGRQVTVQ3FALLPFTIFDETKSWYFTENNERNCAPCNIQMEDPFPKNVRPHA 1920
1940 INGYIMDTPGLVMAQDQIRWYLLSMGNSNENIHSIHFSGHVFTVRKKEEKVMAALYNLYP 1999
1921 INGYIMDTPGLVMAQDQIRWYLLSMGNSNENIHSIHFSGHVFTVRKKEEKVMAALYNLYP 1980
2000 GVFEVEMLPKAGIWRVBCULGEHLHAGMSTLFLVYNKCOOTPLCMASGHIRDQITAS 2059
1981 GVFEVEMLPKAGIWRVBCULGEHLHAGMSTLFLVYNKCOOTPLCMASGHIRDQITAS 2040
2060 GOYGOWAPKLARLHYSGSINAWSTKEPFSWIKVDLLAPMI1HG1KTOGARQKFSLSYISQ 2119
2041 GOYGOWAPKLARLHYSGSINAWSTKEPFSWIKVDLLAPMI1HG1KTOGARQKFSLSYISQ 2100
2120 FIIMTSLDCKKQWTVRGNSGTGLMVFFGNVDSSG1KHNI1FNP2I1ARIYLR1LHPHTYSIRS 2179
2101 FIIMTSLDCKKQWTVRGNSGTGLMVFFGNVDSSG1KHNI1FNP2I1ARIYLR1LHPHTYSIRS 2160
2180 TLRMELMGCDLNSC5MPLGMSKALSDAQITASSYFTNMFKATWSKARLH1QGRSNAR 2239
2161 TLRMELMGCDLNSC5MPLGMSKALSDAQITASSYFTNMFKATWSKARLH1QGRSNAR 2220
2240 POVNNPKEWLQVDFOKTKMKTGVTTTQGVKSLT5MYVKEFTL1SS5ODGHOWTLPFQNGKV 2299
2221 POVNNPKEWLQVDFOKTKMKTGVTTTQGVKSLT5MYVKEFTL1SS5ODGHOWTLPFQNGKV 2280
2300 KVFQGNQDSFTPVVNSLDPPLLTRYLRIHPQSVWHQ1ALRKEVILGCEAQDLY 2351
2281 KVFQGNQDSFTPVVNSLDPPLLTRYLRIHPQSVWHQ1ALRKEVILGCEAQDLY 2332

```

; Sequence 2, Application US/09315179
; Patent No. 6376463
; GENERAL INFORMATION:
; APPLICANT: Lollar, John S
; TITLE OF INVENTION: Modified Factor VII:
; FILE REFERENCE: 75-95H
; CURRENT APPLICATION NUMBER: US/09/315,179
; CURRENT FILING DATE: 1999-05-20
; EARLIER APPLICATION NUMBER: U.S. 09/037,601
; EARLIER FILING DATE: 1998-03-10
; EARLIER APPLICATION NUMBER: U.S. 08/670,707
; EARLIER FILING DATE: 1996-06-26
; EARLIER APPLICATION NUMBER: PCT/US97/11155
; EARLIER FILING DATE: 1997-06-26
; EARLIER APPLICATION NUMBER: PCT/US94/13200
; EARLIER FILING DATE: 1994-11-15
; EARLIER APPLICATION NUMBER: U.S. 08/212,133
; EARLIER FILING DATE: 1994-03-11
; EARLIER APPLICATION NUMBER: U.S. 07/964,004
; NUMBER OF SEQ ID NOS: 40
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 2
; LENGTH: 2332
; TIPS: PRT
; ORGANISM: Homo sapiens
US-09-315-179-2

Query Match          99.0%; Score 12293; DB 4; Length 2332;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 2329; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 20 ATRRYYLGAVELSWDMQSDLGELPVDARPPRPVPSPPFNVSVYKTLFVEFTDHLFN 79
DB 1 ATRRYYLGAVELSWDMQSDLGELPVDARPPRPVPSPPFNVSVYKTLFVEFTVHLEN 60

QY 80 IAKPRPPWMLLGPTIOAEVYTVITLKNMASHPVSLHAGVSVKASEGAEYDDQTSQ 139
DB 61 IAKPRPPWMLLGPTIOAEVYTVITLKNMASHPVSLHAGVSVKASEGAEYDDQTSQ 120

QY 140 REKEDKVPFGSGHTVYVQVLKENGPMASDPICLTYSYLHVDLVKDLNSGLIGALLVCR 199
DB 121 REKEDKVPFGSGHTVYVQVLKENGPMASDPICLTYSYLHVDLVKDLNSGLIGALLVCR 180

QY 200 EGS LAKEKTQTLKFTLLPAVDEGKSHSETKNSLMODRDAASARAWPKHTVANGYNR 259
DB 181 EGS LAKEKTQTLKFTLLPAVDEGKSHSETKNSLMODRDAASARAWPKHTVANGYNR 240

QY 260 SLPGILGCHRSVYVHVIGMGTTPEVHSIFLGHFTFLVRNHRQASLEISPIFTFLTAQTL 319
DB 241 SLPGILGCHRSVYVHVIGMGTTPEVHSIFLGHFTFLVRNHRQASLEISPIFTFLTAQTL 300

QY 320 MDLGQFLLECHSSHQDGMAYVVDSCPERPQLRMKKNBEAEDYDDDLTDSMDVVRFF 379
DB 301 MDLGQFLLECHSSHQDGMAYVVDSCPERPQLRMKKNBEAEDYDDDLTDSMDVVRFF 360

QY 380 DDNSPSTQIRSVAKGPKTWVHYIAAEEEDWDYAPLVAPDDRSYKSQYLNNGPQRIG 439
DB 361 DDNSPSTQIRSVAKGPKTWVHYIAAEEEDWDYAPLVAPDDRSYKSQYLNNGPQRIG 420

QY 440 RYKAVRWMYTDEFTFKTREATOHESGILGPLLYGEVGDITLLIIPKNQASRPYNTYPHGI 499
DB 421 RYKAVRWMYTDEFTFKTREATOHESGILGPLLYGEVGDITLLIIPKNQASRPYNTYPHGI 480

QY 500 TDVRLYRRRLPKGVKHLKDPILPCEIFKYKWTYTVRDGPTKSDPRCLTRYSSFWNME 559
DB 481 TDVRLYRRRLPKGVKHLKDPILPCEIFKYKWTYTVRDGPTKSDPRCLTRYSSFWNME 540

QY 560 RDLASGLIGPLLCYKESVDQRGNQIMSDKRNVLFSVFDENRSWYLTENIQRFPLNPAG 619
DB 541 RDLASGLIGPLLCYKESVDQRGNQIMSDKRNVLFSVFDENRSWYLTENIQRFPLNPAG 600

QY 620 VQLEDPEFQASNMHMSINGYVFDLSQLSVCLHEVAYWYILSIGAQTDFLSVFFSGYTFKH 679

```

Db 1681 DEDNQSPRSFOKTRHYFIAAVERLWDYGMSSSPHLNRNQAQSGSVQFQKVVQFEFTD 1740
Qy 1760 GSFTOPLVGEINELHGLLGPVIRAEVDNIMVTRNQASRSYSYSSLSIYEEDORQA 1819
Db 1741 GSFTOPLVGEINELHGLLGPVIRAEVDNIMVTRNQASRSYSYSSLSIYEEDORQA 1800
Qy 1820 EPRKNFVPNETKTYFWKVQHMAPTKDFCDKAWAYFSDVDLEKDVHSGGLICPLLCHT 1879
Db 1801 EPRKNFVPNETKTYFWKVQHMAPTKDFCDKAWAYFSDVDLEKDVHSGGLICPLLCHT 1860
Qy 1880 NTINPAHGRQVTVQSPALPFTTIDETKSWYFTENMERNCRAPCNTQMEDPTPKENYRPHA 1939
Db 1861 NTINPAHGRQVTVQSPALPFTTIDETKSWYFTENMERNCRAPCNTQMEDPTPKENYRPHA 1920
Qy 1940 INGYIMDTLPLGLVMAQDQIRWVLSMGSNENIHSIHRSGHVFTVRKKEEYKMALYNLYP 1999
Db 1921 INGYINDTLPLGLVMAQDQIRWVLSMGSNENIHSIHRSGHVFTVRKKEEYKMALYNLYP 1980
Qy 2000 GVFETVEMLPKAGIWRVCECLIGEHLHAGMSTLFLVYSNKCOTPLGMAASHIRDFQITAS 2059
Db 1981 GVFETVEMLPKAGIWRVCECLIGEHLHAGMSTLFLVYSNKCOTPLGMAASHIRDFQITAS 2040
Qy 2060 GOYGOWAPKALHSGSINASTKEPFSWIKVDLLAPMIHGIKTQGARQFSSLYISQ 2119
Db 2041 GOYGOWAPKALHSGSINASTKEPFSWIKVDLLAPMIHGIKTQGARQFSSLYISQ 2100
Qy 2120 FIIMYSLDGKQWQYRGNGSTGLMVPFGNVDSGSIKHNFNPPIIARYIRLHPHTYSIRS 2179
Db 2101 FIIMYSLDGKQWQYRGNGSTGLMVPFGNVDSGSIKHNFNPPIIARYIRLHPHTYSIRS 2160
Qy 2180 TLRMELMGCDLNSCSMPLQWESKAISDAQITASSYFTNNFATWSPSKARLHLQGRSNWR 2239
Db 2161 TLRMELMGCDLNSCSMPLQWESKAISDAQITASSYFTNNFATWSPSKARLHLQGRSNWR 2220
Qy 2240 POWNPEKMLQVDPQXTKAVTGTTCGVKSLITSYVVKSLISSODGQWTLFPQNGKY 2299
Db 2221 POWNPEKMLQVDPQXTKAVTGTTCGVKSLITSYVVKSLISSODGQWTLFPQNGKY 2280
Qy 2300 KYFQGNQDSFTPVNSLDPLTRVRIHRPQSWHQAIRMEVLGCEAODLY 2351
Db 2281 KYFQGNQDSFTPVNSLDPLTRVRIHRPQSWHQAIRMEVLGCEAODLY 2332

RESULT 13

US-09-523-656-2
; Sequence 2, Application US/09523656
; Patent No. 6458563
; GENERAL INFORMATION:
; APPLICANT: Lollar S., John
; TITLE OF INVENTION: MODIFIED FACTOR VII
; FILE REFERENCE: 75-951
; CURRENT APPLICATION NUMBER: US/09/523,656
; CURRENT FILING DATE: 2000-03-10
; EARLIER APPLICATION NUMBER: 09/037,601
; EARLIER FILING DATE: 1998-03-10
; EARLIER APPLICATION NUMBER: 08/670,707
; EARLIER FILING DATE: 1996-06-26
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 2332
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-523-656-2

Query Match 99.0%; Score 12293; DB 4; Length 2332;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 2329; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
Qy 20 ATRYVILGAVELSDYMQSDLGELPVDARPPRPVPSFPNTSWYKTLFVEFTDHLFN 79
Db 1 ATRYVILGAVELSDYMQSDLGELPVDARPPRPVPSFPNTSWYKTLFVEFTDHLFN 60

Qy 80 IAKPRPPMGLGTTIOAEVYDTVVITLKNWASHPVSLHVGVSYNKASGAEAYDDOTSQ 139
Db 51 IAKPRPPMGLGTTIOAEVYDTVVITLKNWASHPVSLHVGVSYNKASGAEAYDDOTSQ 120
Qy 140 REKEDDKVPFGSSTYVQVILKENGPMASDPLCTYSYLSHVLDLWKLNSGLIGALLVCR 199
Db 121 REKEDDKVPFGSSTYVQVILKENGPMASDPLCTYSYLSHVLDLWKLNSGLIGALLVCR 180
Qy 200 EGSIAKEKTOTLHKFILLFAVFDGKSMHSETKSLMODRDAASARAWPKMHTVNGYNR 259
Db 181 EGSIAKEKTOTLHKFILLFAVFDGKSMHSETKSLMODRDAASARAWPKMHTVNGYNR 240
Qy 260 SLPLGLIGCHRXSVMHVIGMGTTFEVHSIFLEGFTFLVNRHQASLEISPIITFATAQTL 319
Db 241 SLPLGLIGCHRXSVMHVIGMGTTFEVHSIFLEGFTFLVNRHQASLEISPIITFATAQTL 300
Qy 320 MDLGOPILSCHISSHOHDCMEAYVKVDCPEBPQIRMKNEEAEEDYDDDLTDSMDVVRP 379
Db 301 MDLGOPILSCHISSHOHDCMEAYVKVDCPEBPQIRMKNEEAEEDYDDDLTDSMDVVRP 360
Qy 380 DDNSPSPSTQIRSVAKKHPKTWVHVIAAEBEDWDVAPLVAPDDRSYKSOYLNNPQIRG 439
Db 361 DDNSPSPSTQIRSVAKKHPKTWVHVIAAEBEDWDVAPLVAPDDRSYKSOYLNNPQIRG 420
Qy 440 RYKXKVRFMAYTDETFKTREAIQHESGILGPELLYGEVGDTLIIIFPNQASRPYNIYPHG 499
Db 421 RYKXKVRFMAYTDETFKTREAIQHESGILGPELLYGEVGDTLIIIFPNQASRPYNIYPHG 480
Qy 500 TDVRLYSRRILPKGVKHLKDPILPCEIFKYKWTVTVEDGPTKSDPRCLTYYSFVNM 559
Db 481 TDVRLYSRRILPKGVKHLKDPILPCEIFKYKWTVTVEDGPTKSDPRCLTYYSFVNM 540
Qy 560 RDLASLIGPILLICVKEVDQGNQIMSDKNVILFSVFDENRSMYLTENTQRFLPNPAG 619
Db 541 RDLASLIGPILLICVKEVDQGNQIMSDKNVILFSVFDENRSMYLTENTQRFLPNPAG 600
Qy 620 VQLEDPFQAGNIMESINGYVFDISQLSVCLHEVAYWILSTGAOTDFLSVFSGYTPKH 679
Db 601 VQLEDPFQAGNIMESINGYVFDISQLSVCLHEVAYWILSTGAOTDFLSVFSGYTPKH 660
Qy 680 KMAYEDTTLTFPFGSETVFMENEGLATLGHNSDFENRGMTALLKYSSCDKNIGDYE 739
Db 661 KMAYEDTTLTFPFGSETVFMENEGLATLGHNSDFENRGMTALLKYSSCDKNIGDYE 720
Qy 740 DSYEDISAYLLSKNNAIEPRSPSONSRHPTROKQFNATTIPENDIEKTDPEFAHRTMP 799
Db 721 DSYEDISAYLLSKNNAIEPRSPSONSRHPTROKQFNATTIPENDIEKTDPEFAHRTMP 780
Qy 800 KIONVSSDILLMLRQSPTPHGLS-SDIQEAKYETFPDSDPSGATDSNNSLSEMTHFAPQ 859
Db 781 KIONVSSDILLMLRQSPTPHGLS-SDIQEAKYETFPDSDPSGATDSNNSLSEMTHFAPQ 840
Qy 860 LHSXGDMVFTPESGLOLRNEKLGTTAATLKLDPKVSSTSNLIISTIPSDNLAAGTDN 919
Db 841 LHSXGDMVFTPESGLOLRNEKLGTTAATLKLDPKVSSTSNLIISTIPSDNLAAGTDN 900
Qy 920 TSSLGPPPMFVHYDSQDLDTTLFGKXSSPLTESGGPLSLEENNDKLLSGLMNSQSSW 979
Db 901 TSSLGPPPMFVHYDSQDLDTTLFGKXSSPLTESGGPLSLEENNDKLLSGLMNSQSSW 960
Qy 980 GKNVSTSGRLFKGRAHGPALLTKDNALFKVSIKLTNKTNSNATNRKTHTDGSL 1039
Db 961 GKNVSTSGRLFKGRAHGPALLTKDNALFKVSIKLTNKTNSNATNRKTHTDGSL 1020
Qy 1040 LTENSFSVWQNILEDTPFKVTPLIHDMMLMDKNATLRNLNHNKNTSSKNMVMVQK 1099
Db 1021 LTENSFSVWQNILEDTPFKVTPLIHDMMLMDKNATLRNLNHNKNTSSKNMVMVQK 1080
Qy 1100 KEGPIPPDAQNPMDFEFKMLFLPESARWQIRTHGKNSLNSGQSPKQVLSLGPESKVEG 1159
Db 1081 KEGPIPPDAQNPMDFEFKMLFLPESARWQIRTHGKNSLNSGQSPKQVLSLGPESKVEG 1140
Qy 1160 QNLFSEKKNVVKGEFTKDVGLKEMVFPSSNLFITLNDLHNKNTNQEKIOSEIEK 1219

Db 1141 QNFLSEKXKVVVGKEFTKDVCLKEMVFPSSRNLFITNLNHNNTQKXKIQEETEK 1200
Qy 1220 KETLLCENWVLPQIHTVGTGNFKMKNLFLSTRQNVESYDAGAPVLQDPRSLNDSNR 1279
Db 1201 KETLLCENWVLPQIHTVGTGNFKMKNLFLSTRQNVESYDAGAPVLQDPRSLNDSNR 1260
Qy 1280 TKKHTAHSKKEEENLEGLNQTQKQIIEKACTTRISGNTSQNFVTRQSKRALQKRL 1339
Db 1261 TKKHTAHSKKEEENLEGLNQTQKQIIEKACTTRISGNTSQNFVTRQSKRALQKRL 1320
Qy 1340 PLEETEELKRIIVDDTSTQWSXNMKHLTPSTLTQIDYNEKKGAIQTQPSDCLTRSHSI 1399
Db 1321 PLEETEELKRIIVDDTSTQWSXNMKHLTPSTLTQIDYNEKKGAIQTQPSDCLTRSHSI 1380
Qy 1400 POANRSPPIAKVSSPSPRIPLTRVLFODNSSHLPAASVRKXKDSGVQESSHFLOGAKK 1459
Db 1381 POANRSPPIAKVSSPSPRIPLTRVLFODNSSHLPAASVRKXKDSGVQESSHFLOGAKK 1440
Qy 1460 NNLSLAILTLEMTGQREVSGTSGATKSVTYKKVENTVLPKPDLPKTSKGVELLPKVHI 1519
Db 1441 NNLSLAILTLEMTGQREVSGTSGATKSVTYKKVENTVLPKPDLPKTSKGVELLPKVHI 1500
Qy 1520 YOKDLPTTSTNSGPHLDLVEGSLIQTEGAIKWNEANRPKVPFLRVATESSAKTPSK 1579
Db 1501 YOKDLPTTSTNSGPHLDLVEGSLIQTEGAIKWNEANRPKVPFLRVATESSAKTPSK 1560
Qy 1580 LLDPLAWNHYGTQIPKEBWSQKSPKTAPEKDDTILSNACSNHAIINAINEGQNK 1639
Db 1561 LLDPLAWNHYGTQIPKEBWSQKSPKTAPEKDDTILSNACSNHAIINAINEGQNK 1620
Qy 1640 EIEVTWAKQGRTERLCSQNPVVLKQHQREIRTTILQSQDEIDYDDTISVEMKEDFDIY 1699
Db 1621 EIEVTWAKQGRTERLCSQNPVVLKQHQREIRTTILQSQDEIDYDDTISVEMKEDFDIY 1680
Qy 1700 DEDENQSPRSQKTRHYFIAAVERLWYDGMSSSPHVLNRAQSGSVQPKKVPQOEFTD 1759
Db 1681 DEDENQSPRSQKTRHYFIAAVERLWYDGMSSSPHVLNRAQSGSVQPKKVPQOEFTD 1740
Qy 1760 CSFTQPLVRLGELNEHGLIGPVIRAEVDENIMVTRNOASRPYSFVSSLSIYEEDQROGA 1819
Db 1741 CSFTQPLVRLGELNEHGLIGPVIRAEVDENIMVTRNOASRPYSFVSSLSIYEEDQROGA 1800
Qy 1820 BPRKNFVXPNETKTYFWKQVHMAPTKDEFCDAWAYFSDVDLEKDVHSGLIGPLLVCHT 1879
Db 1801 BPRKNFVXPNETKTYFWKQVHMAPTKDEFCDAWAYFSDVDLEKDVHSGLIGPLLVCHT 1860
Qy 1880 NTLNPAHGRQVTVQEPALPFTTFDETQSWYFTENMERNCRAPCNQMEDPTFKENYRPHA 1939
Db 1861 NTLNPAHGRQVTVQEPALPFTTFDETQSWYFTENMERNCRAPCNQMEDPTFKENYRPHA 1920
Qy 1940 INGYTMDTLPGLVMAQDQRIRWYLLSMGSNENIHFSGHVFTVRKKEEYKXALYNLYP 1999
Db 1921 INGYTMDTLPGLVMAQDQRIRWYLLSMGSNENIHFSGHVFTVRKKEEYKXALYNLYP 1980
Qy 2000 GYFVTEVMLPSKAGIWRVECLIGEHLHAGMSTFLVYNSKQOTPLGMASGHIRDFQITAS 2059
Db 1981 GYFVTEVMLPSKAGIWRVECLIGEHLHAGMSTFLVYNSKQOTPLGMASGHIRDFQITAS 2040
Qy 2060 GYQGWAPKALRHLVSGSINASTKEPFSWIKVDLLAPMIHGIKTQAGARQFSSLYISQ 2119
Db 2041 GYQGWAPKALRHLVSGSINASTKEPFSWIKVDLLAPMIHGIKTQAGARQFSSLYISQ 2100
Qy 2120 FLIYSLDGKKQVTRGNSGTGLVFFFGNVDSG:KHNIENPPIIARYIRLPHTHYSIRS 2179
Db 2101 FLIYSLDGKKQVTRGNSGTGLVFFFGNVDSG:KHNIENPPIIARYIRLPHTHYSIRS 2160
Qy 2180 TURBELMGDLNCSNPLGWSKALSDAQITASSYFTNMFAWSPSKARLHLQGRSNAWR 2239
Db 2161 TURBELMGDLNCSNPLGWSKALSDAQITASSYFTNMFAWSPSKARLHLQGRSNAWR 2220
Qy 2240 POWNNEKEWLQVDFQKTYKVTGTVQKSLITSYVKEFLISSQDGHQWTLFFQNGKV 2299

Db 2221 PQVANNKEWLQVDFQKTKMTKVTGTVTQGVKSLITSYVKEFLISSQDGHQWTLFFQNGKV 2280
Qy 2300 KVFQGNQDSFTFVNVNSLDPPLLTLYLRHPOSWVHOIALRMEVLGCEAODLY 2351
Db 2281 KVFQGNQDSFTFVNVNSLDPPLLTLYLRHPOSWVHOIALRMEVLGCEAODLY 2332
RESULT 14
PCT-US93-03275-4
; Sequence 4, Application PC/TUS9303275
; GENERAL INFORMATION:
; APPLICANT: Lollar, John S.
; APPLICANT: Runge, Marshall S.
; TITLE OF INVENTION: Hybrid Human/Porcine Factor VIII
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Kilpatrick & Cody
; STREET: 1100 Peachtree Street, Suite 2800
; CITY: Atlanta
; STATE: Georgia
; COUNTRY: US
; ZIP: 30309-4530
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US93/03275
; FILING DATE: 19930407
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/864004
; FILING DATE: 07-APR-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Pabst, Patrea L.
; REGISTRATION NUMBER: 31,284
; REFERENCE/DOCKET NUMBER: EMU 106PCT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 404-815-6508
; TELEFAX: 404-815-6555
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2332 amino acids
; TYPE: AMINO ACID
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHEICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: N-terminal
; ORIGINAL SOURCE:
; ORGANISM: Homo sapien
; TISSUE TYPE: Liver cDNA sequence
PCT-US93-03275-4
Query Match 95.0%; Score 12293; DB 5; Length 2332;
Best Local Similarity 95.9%; Pred. No. 0;
Matches 2329; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
Qy 20 ATRYYLGAVELSWDYMOSDLGELPVDARFPFRVPKSPFPNTSVYKTLFVETDHLFN 79
Db 1 ATRYYLGAVELSWDYMOSDLGELPVDARFPFRVPKSPFPNTSVYKTLFVETDHLFN 60
Qy 80 IAKRPPPMWGLGPTIOAEVYDTVTILKNASHPVSLHAGVSVYKASGEAYDDQTSQ 139
Db 61 IAKRPPPMWGLGPTIOAEVYDTVTILKNASHPVSLHAGVSVYKASGEAYDDQTSQ 120
Qy 140 REKEDDKVFPGSGEITYYQVLKENGPMASDPLCLTYSLVSHVDLVKOLNSGLIGALLVCR 199
Db 121 REKEDDKVFPGSGEITYYQVLKENGPMASDPLCLTYSLVSHVDLVKOLNSGLIGALLVCR 180
Qy 200 EGSIAKEXTQTLHKPILLFAVDFDEKSWHSETKXSLMODRDAASARAPRMVTNGYNR 259

181 EGSUAKXTQTLKFLFAVFEDEKSWHSEIKNSLMQDRDASARAWPKHVTNGVNR 240
260 SLPLGILGHRKGVYHVHVGMTTPDVHSTPLEGHTPLVRNHQAQLEISPTFLTAQILL 319
241 SLPLGILGHRKGVYHVHVGMTTPDVHSTPLEGHTPLVRNHQAQLEISPTFLTAQILL 300
320 MDLGFLLLSCH:SSHOHDGMEAYVVDSCPEBPQLRMKNBEAEYDDDLTDSEMDVVRF 379
301 MDLGFLLLSCH:SSHOHDGMEAYVVDSCPEBPQLRMKNBEAEYDDDLTDSEMDVVRF 360
380 DDNSPSPFIQIRSAVAKHPKTVHVIABEEDWDYAPLAPDDRYSYKSQYLNNGPQRIG 439
361 DDNSPSPFIQIRSAVAKHPKTVHVIABEEDWDYAPLAPDDRYSYKSQYLNNGPQRIG 420
440 RYKVKVRNAYTDEFTKTRATQHSGGILGPLLYGEVGDILLIIFKNQASRPYNTYPHGI 499
421 RYKVKVRNAYTDEFTKTRATQHSGGILGPLLYGEVGDILLIIFKNQASRPYNTYPHGI 480
500 TDVRPLYSRRLPKGVKHLKDFPILPGEIPKYKWTVTVEDGPTKSDPRCLTRYSSFVAME 559
481 TDVRPLYSRRLPKGVKHLKDFPILPGEIPKYKWTVTVEDGPTKSDPRCLTRYSSFVAME 540
560 RDLASGLIGPLIICYKESVDQRGNQIMSDKRNVLIFSVDENRSWYLTENIQRFLENPAG 619
541 RDLASGLIGPLIICYKESVDQRGNQIMSDKRNVLIFSVDENRSWYLTENIQRFLENPAG 600
620 VOLEDPEFOASIMHSINGYVDSQLSVCLHEVAYWYILSICAOTDLSVFFSGYTRGH 579
601 VOLEDPEFOASIMHSINGYVDSQLSVCLHEVAYWYILSICAOTDLSVFFSGYTRGH 660
680 KMVYEDTLTLFFSGETVFMSENPCMLTILGCHNSDFRNRGTMALLKVSSCDKNTGDYVE 739
661 KMVYEDTLTLFFSGETVFMSENPCMLTILGCHNSDFRNRGTMALLKVSSCDKNTGDYVE 720
740 DSYEDISAYLLSKNNAIEPRFSQNSRHPSTRQKQFNATTIPENDIEKTDPMFAHRTMP 799
721 DSYEDISAYLLSKNNAIEPRFSQNSRHPSTRQKQFNATTIPENDIEKTDPMFAHRTMP 780
800 KIONVSSDPLMLLQSPFPHGLJSLDQEAKEYTFSDPSFGAIDSNNSLSBMTHERPQ 859
781 KIONVSSDPLMLLQSPFPHGLJSLDQEAKEYTFSDPSFGAIDSNNSLSBMTHERPQ 840
860 LHSXGDMVFTPSGILRLNEXLGTAAELKLPFKVSSSTNNLITSTIPSNLAAGTDN 919
841 LHSXGDMVFTPSGILRLNEXLGTAAELKLPFKVSSSTNNLITSTIPSNLAAGTDN 900
920 TSSLGPPSPNVHYDSQDITLIFGKKSPLTESGGPI:SLSEENNDKLLBSGLMNSQESSW 979
901 TSSLGPPSPNVHYDSQDITLIFGKKSPLTESGGPI:SLSEENNDKLLBSGLMNSQESSW 960
980 GKNVSTSGRLFKGKRAHGPALLTKDNALFKVSI:SLKTKNTKSNNSATNRKTHIDGPSL 1039
961 GKNVSTSGRLFKGKRAHGPALLTKDNALFKVSI:SLKTKNTKSNNSATNRKTHIDGPSL 1020
1040 LIENSPSVWQNLILESDETEPKVTPLTHRMLMDKNATALRLAHMSNKTSSKNMVMVQOK 1099
1021 LIENSPSVWQNLILESDETEPKVTPLTHRMLMDKNATALRLAHMSNKTSSKNMVMVQOK 1080
1100 KEGPIPPDQONPMSFFKMLFPESARWQTHGKNSLNSGGQSPKQVLSIGPEKSVEG 1159
1081 KEGPIPPDQONPMSFFKMLFPESARWQTHGKNSLNSGGQSPKQVLSIGPEKSVEG 1140
1160 QNPLSEKNVWVGKGFTHDVGLEKMPVFPSSNRLFTLNLDNLHNNTHNOKKIOBEIEK 1219
1141 QNPLSEKNVWVGKGFTHDVGLEKMPVFPSSNRLFTLNLDNLHNNTHNOKKIOBEIEK 1200
1220 KETLIQENVVLPQIHTVTGTQKFMKNLFLASTRQNVGSDYAGAYAPVLQDFRSLNDSTNR 1279
1201 KETLIQENVVLPQIHTVTGTQKFMKNLFLASTRQNVGSDYAGAYAPVLQDFRSLNDSTNR 1260
1280 TKKHTAHFSGKGEENLEGNGQTKOIVEKYACTTRESNTSQONFVTOBSKRALKQFRL 1339

1261 TKKHTAHFSGKGEENLEGNGQTKOIVEKYACTTRESNTSQONFVTOBSKRALKQFRL 1320
1340 PLEETELEKRIIVDDTSTOWSKNMKHLTPSLTLOIDYNEKEKGAITQSPDCLTRSHSI 1399
1321 PLEETELEKRIIVDDTSTOWSKNMKHLTPSLTLOIDYNEKEKGAITQSPDCLTRSHSI 1380
1400 PQANRSLPIAKVSSFPISRIRIYILTRVLFQNSSHLPAASRYRKDQSGVQESSHFLQGAKK 1459
1381 PQANRSLPIAKVSSFPISRIRIYILTRVLFQNSSHLPAASRYRKDQSGVQESSHFLQGAKK 1440
1460 NNLSLAILTLEMTGDQREVGLSATSNSVTYKXVENTVLPKPLDPTSGKVLELLPKVHI 1519
1441 NNLSLAILTLEMTGDQREVGLSATSNSVTYKXVENTVLPKPLDPTSGKVLELLPKVHI 1500
1520 YOKLFPPTETGNGSPCHLDLVEGSLLOCTEGAIKWNANRPGKVPFLRVATESAKTPSK 1579
1501 YOKLFPPTETGNGSPCHLDLVEGSLLOCTEGAIKWNANRPGKVPFLRVATESAKTPSK 1560
1580 LLDPLANDNHVGTQIPKEEWSQEKSPKTAFFKKKDTILSNACBSNHAI:AINEGQNK 1639
1561 LLDPLANDNHVGTQIPKEEWSQEKSPKTAFFKKKDTILSNACBSNHAI:AINEGQNK 1620
1640 BIEVTWAKQGRTERLCSQNPVLPKHQREITRTYILQSQOBSIDYDDTTISVEMKKEDFDIY 1699
1621 BIEVTWAKQGRTERLCSQNPVLPKHQREITRTYILQSQOBSIDYDDTTISVEMKKEDFDIY 1680
1700 DEDENQSPRSFQKTRHYFIAAVERLDYGVSSSPHVLNRQAQSGVPOFKKVVQEFBD 1759
1681 DEDENQSPRSFQKTRHYFIAAVERLDYGVSSSPHVLNRQAQSGVPOFKKVVQEFBD 1740
1760 GSFTOPLYRGELNEHLGLGPYIRAEVEDNIMVTFRNOASRPYSFYSSLSYEEDQROGA 1819
1741 GSFTOPLYRGELNEHLGLGPYIRAEVEDNIMVTFRNOASRPYSFYSSLSYEEDQROGA 1800
1820 EPRKNVPKNETKIYFWKVQHHMAPTDEFPCKAWAYESDVLDEKDVHSGLIGPLLVCHT 1879
1801 EPRKNVPKNETKIYFWKVQHHMAPTDEFPCKAWAYESDVLDEKDVHSGLIGPLLVCHT 1860
1880 NTLNPAHQRTVQBEFALLFTIFDETKSWYFTENNERNCRAPCNQIOMEDPTFKENYRHA 1939
1861 NTLNPAHQRTVQBEFALLFTIFDETKSWYFTENNERNCRAPCNQIOMEDPTFKENYRHA 1920
1940 INGYIMDTLPGLVNAQDQIRWYLLSMGSENHISIHFSGHVFTYRKEEYKMALYNLYP 1999
1921 INGYIMDTLPGLVNAQDQIRWYLLSMGSENHISIHFSGHVFTYRKEEYKMALYNLYP 1980
2000 GYFETVEMLPKAGIWRVECLIGEHLAGMSTLFLVYSNKQOTPLGMASGHIRDQITAS 2059
1981 GYFETVEMLPKAGIWRVECLIGEHLAGMSTLFLVYSNKQOTPLGMASGHIRDQITAS 2040
2060 GQYGWAPKLAFLYSGSINAWSTKEPFSWIKVDLLAPMIHGIKTOGARQKFSLSYISQ 2119
2041 GQYGWAPKLAFLYSGSINAWSTKEPFSWIKVDLLAPMIHGIKTOGARQKFSLSYISQ 2100
2120 FIIMYSLDGKKWQTVRGNSTGLMVFVGNVDSGSIKHNINFPPIIARIYIRLHPHTYSIRS 2179
2101 FIIMYSLDGKKWQTVRGNSTGLMVFVGNVDSGSIKHNINFPPIIARIYIRLHPHTYSIRS 2160
2180 TLRMELMCDLNSCSMPLGMESKASDAQITASSYFTNMFEATWSPSKARLHLQGSNAWR 2239
2161 TLRMELMCDLNSCSMPLGMESKASDAQITASSYFTNMFEATWSPSKARLHLQGSNAWR 2220
2240 POWNPKBW:QVDFOKTMKVTVGVTQGVKSLTSMYVKEFLTSSQDGHQWTLFFQNGKV 2299
2221 POWNPKBW:QVDFOKTMKVTVGVTQGVKSLTSMYVKEFLTSSQDGHQWTLFFQNGKV 2280
2300 KVFQNGQSDFTPVVNSLDPPILLTRYLRTHPSQSVHQAIALRMEVLGCEAQDLY 2351
2281 KVFQNGQSDFTPVVNSLDPPILLTRYLRTHPSQSVHQAIALRMEVLGCEAQDLY 2332


```
; Sequence 2, Application PC/TUS9413200
; GENERAL INFORMATION:
; APPLICANT: Emory University
; TITLE OF INVENTION: Hybrid Human/Animal Factor VIII
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Kilpatrick & Cody
; STREET: 1100 Peachtree Street, Suite 2800
; CITY: Atlanta
; STATE: Georgia
; COUNTRY: US
; ZIP: 30309
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US94/13200
; FILING DATE: 15-NOV-1994
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Fabst, Patricia L.
; REGISTRATION NUMBER: 31,284
; REFERENCE/DOCKET NUMBER: EMU106CIP (2)
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 404-815-6508
; TELEFAX: 404-815-6555
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2332 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHEtical: YES
; ANTI-SENSE: NO
; FRAGMENT TYPE: N-terminal
; ORIGINAL SOURCE:
; ORGANISM: Homo sapien
; TISSUE TYPE: Liver cDNA sequence
; PCT-US94-13200-2

Query Match 99.0%; Score 12293; DB 5; Length 2332;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 2329; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 20 ATRRYILGAVELSDWYMQSDLGELPVDARPPRPVPRKSPFNTSVVYKTLFVEFTDHLFN 79
DB 1 ATRRYILGAVELSDWYMQSDLGELPVDARPPRPVPRKSPFNTSVVYKTLFVEFTVHLFN 60

QY 80 IAKPRPPWMLLGPTIOAEVDTVVTTLKNMASHPVSLHAGVSVYKASGEAEYDDOTSQ 139
DB 61 IAKPRPPWMLLGPTIOAEVDTVVTTLKNMASHPVSLHAGVSVYKASGEAEYDDOTSQ 120

QY 140 REKEDDKVPPGGSHYVWQVLKENGEMASDPLCLTYSLSHVDLVKDLNSGLIGALLVCR 199
DB 121 REKEDDKVPPGGSHYVWQVLKENGEMASDPLCLTYSLSHVDLVKDLNSGLIGALLVCR 180

QY 200 EGS LAKEKTQTLHKFILLFAVFDGKSWSEITKNSLMQDRDAASAPAWPKMETVNGYNR 259
DB 181 EGS LAKEKTQTLHKFILLFAVFDGKSWSEITKNSLMQDRDAASAPAWPKMETVNGYNR 240

QY 260 SUPGLIGCHRKSVYWHVIGWTTPEVHSIFLEGHTFLVNRHQASLEISPIFLTAQTL 319
DB 241 SUPGLIGCHRKSVYWHVIGWTTPEVHSIFLEGHTFLVNRHQASLEISPIFLTAQTL 300

QY 320 MDLQGLLCHLSSHQDGMVAVKVYDSCPEPQLRMKNNEAEYDDDLTDSEMDVVR 379
DB 301 MDLQGLLCHLSSHQDGMVAVKVYDSCPEPQLRMKNNEAEYDDDLTDSEMDVVR 360

QY 380 DDNSPSPTQIRSAVKHPTKTVHVI AAEEDWDYAPLV AADDRSYKQYLNNGPQIRG 439
DB 1441 NNLSLAITLLEMTGQREVGSLGTSATNSVTYKAKVENTVLPKDPKPTSGKVLLPKVHI 1500
```

Qy	1520	YQKDLPFTTSSNGSPGHLDIVBESGLIQGTGAIKWNENRPGKVPPLRVATESSAKTPSK	1579
Db	1501	YQKDLPFTTSSNGSPGHLDIVBESGLIQGTGAIKWNENRPGKVPPLRVATESSAKTPSK	1560
Qy	1580	LDLPLAWDNHYGQIIPKEBWKSEKPEKTAFKKOTILSLNACBNHAIALNEGQNP	1639
Db	1561	LDLPLAWDNHYGQIIPKEBWKSEKPEKTAFKKOTILSLNACBNHAIALNEGQNP	1620
Qy	1640	EIEVTWAKQRTBERLCSNPVPLKRQHRETRITTLQSDQEEIDYDITI SVMKKEDFDIY	1699
Db	1621	EIEVTWAKQRTBERLCSNPVPLKRQHRETRITTLQSDQEEIDYDITI SVMKKEDFDIY	1680
Qy	1700	DEBENQSPASFOKKTTHYFTIAAVERLWDYGMSSSPHVLNRAQSGSVPOKKVVOBFTD	1759
Db	1681	DEBENQSPASFOKKTTHYFTIAAVERLWDYGMSSSPHVLNRAQSGSVPOKKVVOBFTD	1740
Qy	1760	GSTFQPLRGELNEHGLLGPYRAEVEDNIMWTFRNQASRPYSFYSSLIISVEDORQGA	1819
Db	1741	GSTFQPLRGELNEHGLLGPYRAEVEDNIMWTFRNQASRPYSFYSSLIISVEDORQGA	1800
Qy	1820	EPRXNFWKPNETKTYFWKVQHHMAPTQDEDFCKAWAYFSDVDLEKDVHSGLIPLJVCHT	1879
Db	1801	EPRXNFWKPNETKTYFWKVQHHMAPTQDEDFCKAWAYFSDVDLEKDVHSGLIPLJVCHT	1860
Qy	1880	NTLNPAHGRVTVQBFALPFTIIEDEKSWYFTENMERNCRAPCNIOWEPTPKENYRFFIA	1939
Db	1861	NTLNPAHGRVTVQBFALPFTIIEDEKSWYFTENMERNCRAPCNIOWEPTPKENYRFFIA	1920
Qy	1940	INGYIMDTLPLGVMAQDORIRWYLLSGNSNENIHSIHFSGEVFTVKKEBYKXVALNLYP	1999
Db	1921	INGYIMDTLPLGVMAQDORIRWYLLSGNSNENIHSIHFSGEVFTVKKEBYKXVALNLYP	1980
Qy	2000	GVFETVEMPLBSKAGIWRVECLIEGHLHAGMSTTLFVYSNKKQTPFLGMASGHIRDFOITAS	2059
Db	1981	GVFETVEMPLBSKAGIWRVECLIEGHLHAGMSTTLFVYSNKKQTPFLGMASGHIRDFOITAS	2040
Qy	2060	GOYGOWAPKLARLHYSGINAWTKBPFMSIKVDLLAPYIIGHCKTOGARQKESSELYISQ	2119
Db	2041	GOYGOWAPKLARLHYSGINAWTKBPFMSIKVDLLAPYIIGHCKTOGARQKESSELYISQ	2100
Qy	2120	FIIMYSLDGKKWQTYRGNSTGTLMVFFNGYDSSGIKHINFPPIIARYIRLHPHTYSIRS	2179
Db	2101	FIIMYSLDGKKWQTYRGNSTGTLMVFFNGYDSSGIKHINFPPIIARYIRLHPHTYSIRS	2160
Qy	2180	TLRMEJMGCDLNSCNPGLWESKAI SDAOITASSYFTNMFWATWSPSKARLHLOGRSNAMR	2239
Db	2161	TLRMEJMGCDLNSCNPGLWESKAI SDAOITASSYFTNMFWATWSPSKARLHLOGRSNAMR	2220
Qy	2240	PQVNNPKWEIQLVDFQTKMKVTGVTTQGVKSILLTSMYVKEFLISSSDGCHQWTLFFQNGKV	2299
Db	2221	PQVNNPKWEIQLVDFQTKMKVTGVTTQGVKSILLTSMYVKEFLISSSDGCHQWTLFFQNGKV	2280
Qy	2300	KVFQGNODSTPPVNSLDPBLLTRYLRBIHPQSWWHQJIALNEVLGECAQDLY	2351
Db	2281	KVFQGNODSTPPVNSLDPBLLTRYLRBIHPQSWWHQJIALNEVLGECAQDLY	2332

Search completed: April 13, 2004, 14:08:23
Job time : 31.5 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 13, 2004, 14:08:32 ; Search time 61.5 Seconds
(without alignments)
10051.498 Million cell updates/sec

Title: NP000123-328
Perfect score: 12416
Sequence: 1 MQIELSTCFCLLRFCFSFSA.....VVHQIALRMEVLGCEAQLDY 2351

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1073127 seqs, 262937947 residues

Total number of hits satisfying chosen parameters: 1073127

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications AA:
1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
3: /cgn2_6/ptodata/2/pubpaa/US05_NEW_PUB.pep.*
4: /cgn2_6/ptodata/2/pubpaa/US05_PUBCOMB.pep.*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep.*
10: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep.*
11: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep.*
12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
13: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep.*
14: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep.*
15: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep.*
16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	12410	100.0	2351	14	US-10-132-829-4
2	12410	100.0	2351	14	US-10-172-712-27
3	12410	100.0	2351	14	US-10-133-907-4
4	12401	99.9	2351	12	US-10-411-037-30
5	12401	99.9	2351	12	US-10-411-026-30
6	12306	99.1	2332	15	US-10-360-101-229
7	12293	99.0	2332	9	US-09-957-641-2
8	12293	99.0	2332	14	US-10-187-319-2
9	12293	99.0	2332	14	US-10-131-510A-2
10	12293	99.0	2332	15	US-10-445-235-2
11	12293	99.0	2332	16	US-10-239-498A-2
12	8827	71.1	2319	14	US-10-187-319-5
13	8827	71.1	2319	14	US-10-131-510A-6
14	8668	65.0	2133	14	US-10-187-319-37
15	8668	65.0	2133	14	US-10-131-510A-37

16	7404	59.6	1471	12	US-10-681-970-2	Sequence 2, Appli
17	7404	59.6	1471	13	US-10-095-718-2	Sequence 2, Appli
18	7305	58.8	1459	13	US-10-239-498A-13	Sequence 13, Appl
19	7295	58.8	1459	16	US-10-239-498A-15	Sequence 15, Appl
20	7292	58.7	1459	16	US-10-239-498A-4	Sequence 4, Appli
21	7219	58.1	1438	13	US-10-006-091-1	Sequence 1, Appli
22	7219	58.1	1438	13	US-10-047-257-1	Sequence 1, Appli
23	7219	58.1	1438	14	US-10-225-900-1	Sequence 1, Appli
24	6264	50.5	1431	12	US-10-681-970-4	Sequence 4, Appli
25	6264	50.5	1431	13	US-10-095-718-4	Sequence 4, Appli
26	6114	49.2	1443	14	US-10-187-319-39	Sequence 39, Appl
27	6114	49.2	1443	14	US-10-131-510A-39	Sequence 39, Appl
28	2824.5	22.7	2196	14	US-10-115-563-14	Sequence 14, Appl
29	2822.5	22.7	2196	15	US-10-160-503-234	Sequence 234, App
30	2822.5	22.7	2196	15	US-10-360-101-259	Sequence 259, App
31	1726	13.9	368	4	US-10-172-712-31	Sequence 31, Appl
32	1726	13.9	368	4	US-10-187-319-4	Sequence 4, Appli
33	1339.5	10.8	1160	12	US-10-131-510A-4	Sequence 4, Appli
34	1339.5	10.8	1160	12	US-10-147-493-234	Sequence 234, App
35	1339.5	10.8	1160	12	US-10-145-127-234	Sequence 234, App
36	1339.5	10.8	1160	12	US-10-160-503-234	Sequence 234, App
37	1339.5	10.8	1160	12	US-10-143-118-234	Sequence 234, App
38	1339.5	10.8	1160	12	US-10-144-993-234	Sequence 234, App
39	1339.5	10.8	1160	12	US-10-158-787-234	Sequence 234, App
40	1339.5	10.8	1160	12	US-10-140-024-234	Sequence 234, App
41	1339.5	10.8	1160	14	US-10-121-049-234	Sequence 234, App
42	1339.5	10.8	1160	14	US-10-123-904-234	Sequence 234, App
43	1339.5	10.8	1160	14	US-10-140-470-234	Sequence 234, App
44	1339.5	10.8	1160	14	US-10-175-746-234	Sequence 234, App
45	1339.5	10.8	1160	14	US-10-176-918-234	Sequence 234, App

ALIGNMENTS

RESULT 1
US-10-132-829-4
; Sequence 4, Application US/10132829
; Publication No. US20030044982A1
; GENERAL INFORMATION:
; APPLICANT: Chien, Kenneth R
; APPLICANT: Hoshitima, Masahiko
; TITLE OF INVENTION: Method to treat hemophilia by hepatic gene transfer of Factor VII:
; FILE OF INVENTION: with vesicle vector
; FILE REFERENCE: 6627-PAL170
; CURRENT APPLICATION NUMBER: US/10132,829
; CURRENT FILING DATE: 2002-04-25
; PRIOR APPLICATION NUMBER: 60/286,314
; PRIOR FILING DATE: 2001-04-25
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 4
; LENGTH: 2351
; TYPE: PRI
; ORGANISM: Homo sapiens
US-10-132-829-4

Query Match	100.0%	Score	12410;	DB	14;	Length	2351;
Best Local Similarity	100.0%	Pred. No.	0;				
Matches	2350;	Conservative	0;	Mismatches	1;	Indels	0;
Gaps	0;						
QY	1	MQIELSTCFCLLRFCFSFSA	RYRYYLGAVELSWDYMOSDLGELPVDARFPPRVPKSPFN	60			
Db	1	MQIELSTCFCLLRFCFSFSA	RYRYYLGAVELSWDYMOSDLGELPVDARFPPRVPKSPFN	60			
QY	61	TSVYKTKTLFVEFTDHLFNIAKPPPMGLLGPTIOAEVYDVTITLKNMASHPVSLHAV	120				
Db	61	TSVYKTKTLFVEFTDHLFNIAKPPPMGLLGPTIOAEVYDVTITLKNMASHPVSLHAV	120				
QY	121	GVSWKASEGAEYDDQTSOREKDDKVPFGGSHYVWQVLKENGPMASDPCLTYSLSH	180				
Db	121	GVSWKASEGAEYDDQTSOREKDDKVPFGGSHYVWQVLKENGPMASDPCLTYSLSH	180				

181 VDLVKDLSGLIGALLVCREGLAKEKQTQTLHKPILLFAVDEGKSWHSEKNSLMQDRD 240
181 VDLVKDLSGLIGALLVCREGLAKEKQTQTLHKPILLFAVDEGKSWHSEKNSLMQDRD 240
241 AASARAWPMHTVNGVYNSRPLGLGCHRSKYVYHVI GMGTTPEVHSLFLEHPTLVNRH 300
241 AASARAWPMHTVNGVYNSRPLGLGCHRSKYVYHVI GMGTTPEVHSLFLEHPTLVNRH 300
301 ROASLEI SPITLACTLLMDLGOFLSCHEISHOHGDGMEAVVKYDSCPEEPOLRMKNE 360
301 ROASLEI SPITLACTLLMDLGOFLSCHEISHOHGDGMEAVVKYDSCPEEPOLRMKNE 360
36 EADYDDDDITDSEMDVVRPDDDNSSPSFIQIRSVAKKHGPKTWVHYIAABEEDWDYAPLVIA 420
361 EADYDDDDITDSEMDVVRPDDDNSSPSFIQIRSVAKKHGPKTWVHYIAABEEDWDYAPLVIA 420
421 PDRSYKSOYLNGPQIRGRKYKVRPMAYTDETFKTREAIHESGILGPLLYGEGDTL 480
421 PDRSYKSOYLNGPQIRGRKYKVRPMAYTDETFKTREAIHESGILGPLLYGEGDTL 480
481 LIIFKNOASRPYNIYPHGITDVRPLYSRRLPKGVKHLKDPFIPGLPGEIFKYKMTVTVEDGP 540
481 LIIFKNOASRPYNIYPHGITDVRPLYSRRLPKGVKHLKDPFIPGLPGEIFKYKMTVTVEDGP 540
541 TKSDPRCLTRYSSVFVNMERDILASGLIGPLLYCYKESVDQRGNQIMSDKRNVLFSVDE 600
541 TKSDPRCLTRYSSVFVNMERDILASGLIGPLLYCYKESVDQRGNQIMSDKRNVLFSVDE 600
601 NRSWYLTEINQIFLNPAGVQLEDPEFOASNIMHSINGVYVDSQLSVCLHEVAVWYILS 660
601 NRSWYLTEINQIFLNPAGVQLEDPEFOASNIMHSINGVYVDSQLSVCLHEVAVWYILS 660
661 IGAQTDFLSVFSGVTFKHKMYVEDTLTLPFSGETVFMSMNPGLWILGCHNSDFNRG 720
661 IGAQTDFLSVFSGVTFKHKMYVEDTLTLPFSGETVFMSMNPGLWILGCHNSDFNRG 720
721 MTALIKVSSCDKNTGYVEDSDVEDISAYLLSKNAIEPRSFSONSRHPSTROKQFNATTI 780
721 MTALIKVSSCDKNTGYVEDSDVEDISAYLLSKNAIEPRSFSONSRHPSTROKQFNATTI 780
781 PENDIEKTDPMFAHRTPMBKIQNVSSDLMLLRQSPTEHG:SLSDLOBAKYETFSDDPS 840
781 PENDIEKTDPMFAHRTPMBKIQNVSSDLMLLRQSPTEHG:SLSDLOBAKYETFSDDPS 840
841 PGADSNNSLSLSEMTFRPQLHHSQDMVFTPESGQLRLNEKLGTTAATLKKLDFXVSST 900
841 PGADSNNSLSLSEMTFRPQLHHSQDMVFTPESGQLRLNEKLGTTAATLKKLDFXVSST 900
901 SNNLISTIPSDNLAACTDNTSSLGPPSMVHYDSQLDITLPGKSSPLTESGGPLSLSEE 960
901 SNNLISTIPSDNLAACTDNTSSLGPPSMVHYDSQLDITLPGKSSPLTESGGPLSLSEE 960
961 NNDKLLSGLMNSQESSWGKNVSSSTESGRLFKGKAHGFALLTKDNALFKYSISLLKTN 1020
961 NNDKLLSGLMNSQESSWGKNVSSSTESGRLFKGKAHGFALLTKDNALFKYSISLLKTN 1020
1021 KTSNNSATNRKTHIDGPSLLIENSPPVQNLIESDTEFKKVTPLIHDRMLMDKNAALRL 1080
1021 KTSNNSATNRKTHIDGPSLLIENSPPVQNLIESDTEFKKVTPLIHDRMLMDKNAALRL 1080
1081 NMSNKTTSKKNMNTVOQKEGPIPPDAQNPDMSFEKMLFLPESARWIORTHCKNSLNSG 1140
1081 NMSNKTTSKKNMNTVOQKEGPIPPDAQNPDMSFEKMLFLPESARWIORTHCKNSLNSG 1140
1141 QGSPKQOLVSLGPEKSVGEGQFLSBNKVVVGKEFTKDVGLKEMVPPSSRNFLTNLDN 1200
1141 QGSPKQOLVSLGPEKSVGEGQFLSBNKVVVGKEFTKDVGLKEMVPPSSRNFLTNLDN 1200
1201 LHENNTNHOEKKIOEBLEKXETLIQSNVLPQ:HTVTGTRKFMKNLFLJSTRQNVEGSYD 1260
1201 LHENNTNHOEKKIOEBLEKXETLIQSNVLPQ:HTVTGTRKFMKNLFLJSTRQNVEGSYD 1260
1261 GAYAPVLODFSLNDSNTRTKKTAHFSKKEEENLEGLGNQTKQIVEKYACTTRISPN 1320

1261 GAYAPVLODFSLNDSNTRTKKTAHFSKKEEENLEGLGNQTKQIVEKYACTTRISPN 1320
1321 SQONFVTVORSKRALKQFELPLEETELBKRIIVDDTSTOWSNKMKHLTPSTLTQIDYNEKE 1380
1321 SQONFVTVORSKRALKQFELPLEETELBKRIIVDDTSTOWSNKMKHLTPSTLTQIDYNEKE 1380
1381 KGAITQSPSLDCLTRSHSIPQANRSPPIAKVSSPPSIRPIYLTFRVLFPQDNSSHLPAASY 1440
1381 KGAITQSPSLDCLTRSHSIPQANRSPPIAKVSSPPSIRPIYLTFRVLFPQDNSSHLPAASY 1440
1441 RKXSGVOESSHPFLOGAKKNLSAILTLEMTGQOREVGSIGTSATNSVTYKKVENTVLP 1500
1441 RKXSGVOESSHPFLOGAKKNLSAILTLEMTGQOREVGSIGTSATNSVTYKKVENTVLP 1500
1501 K3DLPKTSKGYELLPKVHIYQDLPFTSTNSGPHLDLVEGSLLOQTGEGAIKWEANRP 1560
1501 K3DLPKTSKGYELLPKVHIYQDLPFTSTNSGPHLDLVEGSLLOQTGEGAIKWEANRP 1560
1561 GKVPFLRVATESSAKTPSKLLDPLAWDNHYGTQIPKEEWKSOEKSPKTAFFKOTILSL 1620
1561 GKVPFLRVATESSAKTPSKLLDPLAWDNHYGTQIPKEEWKSOEKSPKTAFFKOTILSL 1620
1621 NACESNHAIAAINEGONKPEIEVTWAKQGRTERLCSQNPVVKRQRBITRTTLOSDOEB 1680
1621 NACESNHAIAAINEGONKPEIEVTWAKQGRTERLCSQNPVVKRQRBITRTTLOSDOEB 1680
1681 IDYDDTISVENKKEFDIYDEDENQSPRSFOKTRHYFIAAVERLDYCMSSSPHVLNR 1740
1681 IDYDDTISVENKKEFDIYDEDENQSPRSFOKTRHYFIAAVERLDYCMSSSPHVLNR 1740
1741 AQSGSVPOFKXVVOFETDGSFTQPLYRGELNEHGLLGPYIRAEVEONIMVTFNQASR 1800
1741 AQSGSVPOFKXVVOFETDGSFTQPLYRGELNEHGLLGPYIRAEVEONIMVTFNQASR 1800
1801 PYSFYSSLIISVEEDORQGAEPKNEKTYEYKQHEHMAPTKDPPDCKAWAYFSDV 1860
1801 PYSFYSSLIISVEEDORQGAEPKNEKTYEYKQHEHMAPTKDPPDCKAWAYFSDV 1860
1861 DLEKDVHSLGILPLVCHTNTLNPAGHQVTVQEPALFPTTDETKSWYFTENMERNORA 1920
1861 DLEKDVHSLGILPLVCHTNTLNPAGHQVTVQEPALFPTTDETKSWYFTENMERNORA 1920
1921 PCNIQMEDPTPKENYRFAHNGYIMDTLPGVWAOQRIRWYLLSMGNSNENIHSIHFSGH 1980
1921 PCNIQMEDPTPKENYRFAHNGYIMDTLPGVWAOQRIRWYLLSMGNSNENIHSIHFSGH 1980
1981 VFTVRKKEBYKVALNLYPGVFEVEMLPKAGIWRVECLIGEHLHAGMSTLFLVYSNKC 2040
1981 VFTVRKKEBYKVALNLYPGVFEVEMLPKAGIWRVECLIGEHLHAGMSTLFLVYSNKC 2040
2041 QTPLGWASGHIRDPQITASQGYQWAPKLARLHYSGSINAWSTKEFFSWIKVDLLAPMII 2100
2041 QTPLGWASGHIRDPQITASQGYQWAPKLARLHYSGSINAWSTKEFFSWIKVDLLAPMII 2100
2101 HGKIKQAROKSSLYISQPIIIMYSLDCKKQWQTYGNSTGLMWFGNVDSGCIKHNIEN 2160
2101 HGKIKQAROKSSLYISQPIIIMYSLDCKKQWQTYGNSTGLMWFGNVDSGCIKHNIEN 2160
2161 PPIIARYIRLHPTHYSIRSTRLWELMGDLNCSMPLGMESKAI SDAQITASSYTNMFA 2220
2161 PPIIARYIRLHPTHYSIRSTRLWELMGDLNCSMPLGMESKAI SDAQITASSYTNMFA 2220
2221 TWSPSKARLHLQGRSNARPPQVNNPKMLQVDFOKTMKVTVGTTQGVKSLTSMYVKBEPL 2280
2221 TWSPSKARLHLQGRSNARPPQVNNPKMLQVDFOKTMKVTVGTTQGVKSLTSMYVKBEPL 2280
2281 ISSSQDGHQWTLFPQNGKVKYQGNQDSFTPVNSLDPPLLTLYLRHPQSWHQAIALRM 2340
2281 ISSSQDGHQWTLFPQNGKVKYQGNQDSFTPVNSLDPPLLTLYLRHPQSWHQAIALRM 2340
2341 EYLGCEAODLY 2351

Db	2341	EVLGCAQDLY	2351
		RESULT 2	
		US-10-172-712-27	
		Sequence 27, Application US/10172712	
		Publication No. US20030125232A1	
		GENERAL INFORMATION:	
		APPLICANT: GRIFFIN, JOHN H.	
		APPLICANT: GALE, ANDREW J.	
		APPLICANT: GETZOFF, ELIZABETH D.	
		APPLICANT: PELLEQUER, JEAN-LUC	
		TITLE OF INVENTION: STABILIZED PROTEINS WITH ENGINEERED DISULFIDE BONDS	
		FILE REFERENCE: 4198-4001US1	
		CURRENT APPLICATION NUMBER: US/10/172,712	
		PRIOR FILING DATE: 2002-09-30	
		CURRENT FILING DATE: 2002-09-30	
		PRIOR FILING DATE: 2001-06-14	
		NUMBER OF SEQ ID NOS: 32	
		SOFTWARE: PatentIn ver. 2.1	
		SEQ ID NO 27	
		LENGTH: 2351	
		TYPE: PRT	
		ORGANISM: Homo sapiens	
		US-10-172-712-27	
		Query Match 100.0%; Score 12410; DB 14; Length 2351;	
		Best Local Similarity 100.0%; Pred. No. 0;	
		Matches 2350; Conservative 0; Mismatches 1; Indels 0; Gaps 0;	
QY	1	MOEISTCFCLLRFCFSATRRYYLGAVELSWDMQSDLGELPVDARFPFRVPKSPFN	60
Db	1	MOEISTCFCLLRFCFSATRRYYLGAVELSWDMQSDLGELPVDARFPFRVPKSPFN	60
QY	61	TSVWKKCLFVETDELNIAPRPWMLGPTTQAEVVDVTVITLKNMASHPVSLHAV	120
Db	61	TSVWKKCLFVETDELNIAPRPWMLGPTTQAEVVDVTVITLKNMASHPVSLHAV	120
QY	121	GVSYWKASGAEDDQTSQREKDDKVPFGGSHYTVWQVLKENGPMASDPLCLTYSYLSH	180
Db	121	GVSYWKASGAEDDQTSQREKDDKVPFGGSHYTVWQVLKENGPMASDPLCLTYSYLSH	180
QY	181	VDLVKDLNSGLICALLVREGSLAKEKQVTLHKFTLLFAVFDDEKSHSETPKNSLMQDRD	240
Db	181	VDLVKDLNSGLICALLVREGSLAKEKQVTLHKFTLLFAVFDDEKSHSETPKNSLMQDRD	240
QY	241	AASARAWPMHWTVNGVYNSRLPCLIGCHRKSVYWHVIGMTGTPPVHVSIFLEGHTFLVRNH	300
Db	241	AASARAWPMHWTVNGVYNSRLPCLIGCHRKSVYWHVIGMTGTPPVHVSIFLEGHTFLVRNH	300
QY	301	ROASLEISPIITLTAQTLMDLIGQFLLSCHTSSHQHDGMEAYVKVDSCPEBPQLRMKNNE	360
Db	301	ROASLEISPIITLTAQTLMDLIGQFLLSCHTSSHQHDGMEAYVKVDSCPEBPQLRMKNNE	360
QY	361	EAEVDYDDLTDSEMDVVRPDDNSPSFQIRSVAKKHPKTVWHVYIAAPERDWDVAPLVLA	420
Db	361	EAEVDYDDLTDSEMDVVRPDDNSPSFQIRSVAKKHPKTVWHVYIAAPERDWDVAPLVLA	420
QY	421	PDRSYKGOYLANGPQIRGRKYKVRWAYTDETPKTRTREAIOHESGILGPLLYGEVGDTL	480
Db	421	PDRSYKGOYLANGPQIRGRKYKVRWAYTDETPKTRTREAIOHESGILGPLLYGEVGDTL	480
QY	481	LIIFKNAQRPNINYPHGITDVRPLYSRLPKGVKHLKDFILGELPKYKWTVTYVDDGP	540
Db	481	LIIFKNAQRPNINYPHGITDVRPLYSRLPKGVKHLKDFILGELPKYKWTVTYVDDGP	540
QY	541	TKSDPRCLTRYSSFPVNMRDILASGLIGPLLIICYKESVDQRGNQIMSKRNVIILFSVFDE	600
Db	541	TKSDPRCLTRYSSFPVNMRDILASGLIGPLLIICYKESVDQRGNQIMSKRNVIILFSVFDE	600
QY	601	NRSWYLTENIQRLPNPAQVLEDEPFOASNIMHISINGVYFDSQLSVCLHEVAYWILS	660
Db	601	NRSWYLTENIQRLPNPAQVLEDEPFOASNIMHISINGVYFDSQLSVCLHEVAYWILS	660

```
QY 1741 AQSGVPOFKKVVFOEFTDGSFTQPLRGELNEHLGLGPYIRAEVEDNMVTFNQASR 1800
Db 1741 AQSGVPOFKKVVFOEFTDGSFTQPLRGELNEHLGLGPYIRAEVEDNMVTFNQASR 1800
QY 1801 PYSFYSSLSYEDORQAGAPKRNFKVKNETKTYFKVQHVNAPTKDBDPCKAWAYFSDV 1860
Db 1801 PYSFYSSLSYEDORQAGAPKRNFKVKNETKTYFKVQHVNAPTKDBDPCKAWAYFSDV 1860
QY 1861 DAEKDVHSLGELIPLVCHNTNLPAGHROVTVQESALPFTIFDETKSWYFTENMERNCBA 1920
Db 1861 DAEKDVHSLGELIPLVCHNTNLPAGHROVTVQESALPFTIFDETKSWYFTENMERNCBA 1920
QY 1921 PCNIQWEDPTFKENYAFHAINGYIMDTPLGLVMAQDQRIWYLLSMGNSNENHSIFSGH 1980
Db 1921 PCNIQWEDPTFKENYAFHAINGYIMDTPLGLVMAQDQRIWYLLSMGNSNENHSIFSGH 1980
QY 1981 VFTVRKKEBYKVALYNLYPGVFTVEMLPKAGIWEVECLIGEHLHAGNSTFLVYSNKC 2040
Db 1981 VFTVRKKEBYKVALYNLYPGVFTVEMLPKAGIWEVECLIGEHLHAGNSTFLVYSNKC 2040
QY 2041 QTPPLGMAHGHIRDFQITASGOYQWAPKLARLHYSGEINAWSTKBPFSWIKVDLLAPMII 2100
Db 2041 QTPPLGMAHGHIRDFQITASGOYQWAPKLARLHYSGEINAWSTKBPFSWIKVDLLAPMII 2100
QY 2101 HGIKTQARQKSSLYISOFIIMYSLDGKKWQTYRGNSTGTLMVPPGNVDDSGIKHNIFN 2160
Db 2101 HGIKTQARQKSSLYISOFIIMYSLDGKKWQTYRGNSTGTLMVPPGNVDDSGIKHNIFN 2160
QY 2161 PPIIARYIRLPHPHYSIRSTLRMELMGCDLNSCMPLGWSKAISSAOITASSYFTNMFA 2220
Db 2161 PPIIARYIRLPHPHYSIRSTLRMELMGCDLNSCMPLGWSKAISSAOITASSYFTNMFA 2220
QY 2221 TWSPSKARLHLQGRNAMPPOVNNPKWLQVDFQKTKMKTGVTITQGVKSLLTSMYVKEPL 2280
Db 2221 TWSPSKARLHLQGRNAMPPOVNNPKWLQVDFQKTKMKTGVTITQGVKSLLTSMYVKEPL 2280
QY 2281 ISSQDGHQWTLFFQNGKVKVQGNQDSTPPVNSLDPPLLRILRIHQSWVHQIALRM 2340
Db 2281 ISSQDGHQWTLFFQNGKVKVQGNQDSTPPVNSLDPPLLRILRIHQSWVHQIALRM 2340
QY 2341 EVLGCEAQDLY 2351
Db 2341 EVLGCEAQDLY 2351

RESULT 3
US-10-133-907-4
; Sequence 4, Application US/10133907
; Publication No. US2003019523A1
; GENERAL INFORMATION:
; APPLICANT: Chien, Kenneth R
; TITLE OF INVENTION: Method to treat hemophilia by hepatic gene transfer of Factor VIII
; TITLE OF INVENTION: with vesicle vector
; FILE REFERENCE: 6627-Pal170
; CURRENT APPLICATION NUMBER: US/10/133,907
; CURRENT FILING DATE: 2002-04-25
; PRIOR APPLICATION NUMBER: 60/286,314
; PRIOR FILING DATE: 2001-04-25
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 4
; LENGTH: 2351
; TYPE: PRN
; ORGANISM: Homo sapiens
US-10-133-907-4

Query Match 100.0%; Score 12410; DB 14; Length 2351;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2350; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MQIELSTCFCLLRFCFSATRRYYLGAVELSDWYMQSDLGELPYDARPPRPVKSPFFN 60
```

```
Db 1 MQIELSTCFCLLRFCFSATRRYYLGAVELSDWYMQSDLGELPYDARPPRPVKSPFFN 60
QY 61 TSVVYKTLFVSEFTDHLNIAKPRPPWMLLGPITQAEVDTVITLKNMASHPVSLSHAV 120
Db 61 TSVVYKTLFVSEFTDHLNIAKPRPPWMLLGPITQAEVDTVITLKNMASHPVSLSHAV 120
QY 121 GYSYWKASGAEAYDDQTSQREKDDKVPFGSGSHYTVVWLKENGPMASDPCLCLTYSLSH 180
Db 121 GYSYWKASGAEAYDDQTSQREKDDKVPFGSGSHYTVVWLKENGPMASDPCLCLTYSLSH 180
QY 181 VDLVXDLNSGLIGALLVOREGSLAKEKTQTLHKFILLFAVDEGKSWHSETKNSLMQDRD 240
Db 181 VDLVXDLNSGLIGALLVOREGSLAKEKTQTLHKFILLFAVDEGKSWHSETKNSLMQDRD 240
QY 241 AASARAWPMHTVNGYVNRSLPGLIGCHRKSVYWHVIGMGTTPVHSLFLECHTEFLVRNH 300
Db 241 AASARAWPMHTVNGYVNRSLPGLIGCHRKSVYWHVIGMGTTPVHSLFLECHTEFLVRNH 300
QY 301 ROASLEISPIFLTAQTLMDLIGQFLLSCHISSHQHDMGEAYVKYVDSCEPBPQLRMKNE 360
Db 301 ROASLEISPIFLTAQTLMDLIGQFLLSCHISSHQHDMGEAYVKYVDSCEPBPQLRMKNE 360
QY 361 EABDYDDDLTDSMDVVRFDNDSPSFTQIRSAVAKKPKTWVHYIAAEEEDWDYAPLVLA 420
Db 361 EABDYDDDLTDSMDVVRFDNDSPSFTQIRSAVAKKPKTWVHYIAAEEEDWDYAPLVLA 420
QY 421 PDRSYKQYLNNGPQRIGRYKVKVRFMAYTDETPKTEAIOHESGILGLPLYGEGVDTL 480
Db 421 PDRSYKQYLNNGPQRIGRYKVKVRFMAYTDETPKTEAIOHESGILGLPLYGEGVDTL 480
QY 481 LIIFKNQASRPNIYPHGITDVRPIYSRRLPKGVKHLKDPFPLPGSEIKYKWTVTVEGSP 540
Db 481 LIIFKNQASRPNIYPHGITDVRPIYSRRLPKGVKHLKDPFPLPGSEIKYKWTVTVEGSP 540
QY 541 TKSDDPRCLTRYSSVFNVERDLASGLIGPLLCYKESVDQRGNQIMSKNVILFSVDE 600
Db 541 TKSDDPRCLTRYSSVFNVERDLASGLIGPLLCYKESVDQRGNQIMSKNVILFSVDE 600
QY 601 NRSWYLTENIQRFLENPAGVQLEDEFOASNIHMSINGYVFDLSQLSVCLHEVAYWYLS 660
Db 601 NRSWYLTENIQRFLENPAGVQLEDEFOASNIHMSINGYVFDLSQLSVCLHEVAYWYLS 660
QY 661 IGAQTDPLSVFPFGYTFKHVWYEDTLTLPFSGETVFMSENPGIWLGCNDSFRNRG 720
Db 661 IGAQTDPLSVFPFGYTFKHVWYEDTLTLPFSGETVFMSENPGIWLGCNDSFRNRG 720
QY 721 MTALLKVSQCDKNTGDIYEDSVEDISAYLLSKNNAIEPRSPSONSRHSTROKQFNATTI 780
Db 721 MTALLKVSQCDKNTGDIYEDSVEDISAYLLSKNNAIEPRSPSONSRHSTROKQFNATTI 780
QY 781 PENDIEKTDPPFAHRTPMFKIQNVSSSLLMLLRQSPTHGLSLSDLOEAKYETESDPS 840
Db 781 PENDIEKTDPPFAHRTPMFKIQNVSSSLLMLLRQSPTHGLSLSDLOEAKYETESDPS 840
QY 841 PGAISSNLSLSEMTFRPQLHSGGDMVTPESGLQRLNEKLGTTAATLKKLDFKVSST 900
Db 841 PGAISSNLSLSEMTFRPQLHSGGDMVTPESGLQRLNEKLGTTAATLKKLDFKVSST 900
QY 901 SNMLISTIPSDNLAAGTONTSSLGPPSPVHYDSDTLTLPKKSPLTSESGPLSLEE 960
Db 901 SNMLISTIPSDNLAAGTONTSSLGPPSPVHYDSDTLTLPKKSPLTSESGPLSLEE 960
QY 961 NNDSKLLSGLMNSQESSGKRVNSTESGRIFKGRAGHPALLTKDNALFKVSIILLKTIN 1020
Db 961 NNDSKLLSGLMNSQESSGKRVNSTESGRIFKGRAGHPALLTKDNALFKVSIILLKTIN 1020
QY 1021 KTSNNSATNRKTHIDGPSLLIENSVSVMONTLESOTEFKVTPELHDMXMDKNTAURL 1080
Db 1021 KTSNNSATNRKTHIDGPSLLIENSVSVMONTLESOTEFKVTPELHDMXMDKNTAURL 1080
QY 1081 NEMSNTTSSKNMVMQKKEGPIPPDAQNPVSPFFKMLFPESARWQRTHGKXSLNSG 1140
Db 1081 NEMSNTTSSKNMVMQKKEGPIPPDAQNPVSPFFKMLFPESARWQRTHGKXSLNSG 1140
```

1141 QGSPKQVLSLGPKEVGEQNFSEKKNVVGKGEFTKDVGLKEMVPPSSRNLFNTLNL 1200
1141 QGSPKQVLSLGPKEVGEQNFSEKKNVVGKGEFTKDVGLKEMVPPSSRNLFNTLNL 1200
1201 LHENNTNHEKKTQIEIEKKEKTLIQENVVLPIQHTVTGKTNFMKNLFLLSRQNVGSYD 1260
1201 LHENNTNHEKKTQIEIEKKEKTLIQENVVLPIQHTVTGKTNFMKNLFLLSRQNVGSYD 1260
1261 GAYAPVLOPFRSLNDSNTKHTAHSKGBEENLEGLGNQTKQIVKXACTTRISPNPT 1320
1261 GAYAPVLOPFRSLNDSNTKHTAHSKGBEENLEGLGNQTKQIVKXACTTRISPNPT 1320
1321 SOONFVTSKQKALKOPRLPBETELEKRIIVDDTSTOWSKMKHLTPSTLTQIDYNEKE 1380
1321 SOONFVTSKQKALKOPRLPBETELEKRIIVDDTSTOWSKMKHLTPSTLTQIDYNEKE 1380
1381 KGATQSPISDCLTRSHSIPQANRSPPIAKVSSPPSIRPIYLVTRVLFQDNSSHLPAASY 1440
1381 KGATQSPISDCLTRSHSIPQANRSPPIAKVSSPPSIRPIYLVTRVLFQDNSSHLPAASY 1440
1441 RKKDSGVQSSHFLOCAKKNLSLALTLEMTGDQREVGLSATNSVTYKKVENTVLP 1500
1441 RKKDSGVQSSHFLOCAKKNLSLALTLEMTGDQREVGLSATNSVTYKKVENTVLP 1500
1501 KEDLPKTSKVELLPKVHIYQKDLFPETSNQSPGHLDLVEGSLQCTEGALKWNEANRP 1560
1501 KEDLPKTSKVELLPKVHIYQKDLFPETSNQSPGHLDLVEGSLQCTEGALKWNEANRP 1560
1561 GXVPFLRVATSAKTPSKLLDPLANDNHYGTQIPKEWKSEKSPKTAFFKKOTILSL 1620
1561 GXVPFLRVATSAKTPSKLLDPLANDNHYGTQIPKEWKSEKSPKTAFFKKOTILSL 1620
1621 NACESHAAIAANEGONKBPRIEWTAKQGRTELCSQNPVLKSHQREITRTTLOSDDQE 1680
1621 NACESHAAIAANEGONKBPRIEWTAKQGRTELCSQNPVLKSHQREITRTTLOSDDQE 1680
1681 IDYDDTISVEMKKEDFDIYDEDENQSPRSFQKTRHYFIAAVERLWDYGMSSSPHVLNR 1740
1681 IDYDDTISVEMKKEDFDIYDEDENQSPRSFQKTRHYFIAAVERLWDYGMSSSPHVLNR 1740
1741 AQSGVQPQKVVQFQFTDQSTQPIYRGELNEHLLGPGYTRAEVEDNIMVTFRQASR 1800
1741 AQSGVQPQKVVQFQFTDQSTQPIYRGELNEHLLGPGYTRAEVEDNIMVTFRQASR 1800
1801 PYSFYSLLSYBEDQOQAEPRNFVKNETKTYFWKVQHVMAPTKDEDFCKAWAYFSDV 1860
1801 PYSFYSLLSYBEDQOQAEPRNFVKNETKTYFWKVQHVMAPTKDEDFCKAWAYFSDV 1860
1861 DLEKOVHSLIGPLLVCHTNTLNPAGHGRQVTVQEFALFTTIDETKSWYFTENMERNCEA 1920
1861 DLEKOVHSLIGPLLVCHTNTLNPAGHGRQVTVQEFALFTTIDETKSWYFTENMERNCEA 1920
1921 PCNIQWEDTFKENTRFAHNGYIMDTLGLVMAQQRIRWLLSGNSNETHSHFSQH 1980
1921 PCNIQWEDTFKENTRFAHNGYIMDTLGLVMAQQRIRWLLSGNSNETHSHFSQH 1980
1981 VFTVRKKEEYKMAVNLNYPGVETVEMLPKAGIRWEVCLIGHLLHAGMSTLFLVYSNKC 2040
1981 VFTVRKKEEYKMAVNLNYPGVETVEMLPKAGIRWEVCLIGHLLHAGMSTLFLVYSNKC 2040
2041 QTPLGWASGHIRDFOITAGSQGWAPKLARLHYSGSINAWSTKEPFSWKVDLAPMII 2100
2041 QTPLGWASGHIRDFOITAGSQGWAPKLARLHYSGSINAWSTKEPFSWKVDLAPMII 2100
2101 HGIKTGAQKQSSLSVISQFIIMYLDGKKWQTYRGNSTGTLWFFGNVDSSGIKINIEN 2160
2101 HGIKTGAQKQSSLSVISQFIIMYLDGKKWQTYRGNSTGTLWFFGNVDSSGIKINIEN 2160
2161 PPIIARYIILFPCHYISIRGTIRMELMGCDLNSCMPLGMESKAISDAQITASSYFTNMPA 2220
2161 PPIIARYIILFPCHYISIRGTIRMELMGCDLNSCMPLGMESKAISDAQITASSYFTNMPA 2220

2221 TWSPKARLHLQGRNANRPQVNNPKMLQVDFOKTMKVTVTTQGVKSLTSMYKBEFL 2280
2221 TWSPKARLHLQGRNANRPQVNNPKMLQVDFOKTMKVTVTTQGVKSLTSMYKBEFL 2280
2281 ISSSQDGHQWTLFFQNGKVKVFCQNDQSFPPVNSLDPPLLTRYLRIHPQSWVHQIALRM 2340
2281 ISSSQDGHQWTLFFQNGKVKVFCQNDQSFPPVNSLDPPLLTRYLRIHPQSWVHQIALRM 2340
2341 EVLGCQAQDLY 2351
2341 EVLGCQAQDLY 2351

RESULT 4
US-10-411-037-30
; Sequence 30, Application US/10411037
; Publication NO. US20040043446A1
; GENERAL INFORMATION:
; APPLICANT: DeFree, Shawn
; APPLICANT: Zopf, David
; APPLICANT: Bayer, Robert
; APPLICANT: Hakes, David
; APPLICANT: Chen, Xi
; APPLICANT: Bowe, Caryn
; TITLE OF INVENTION: ALPHA GALACTOSIDASE A: REMODELING AND GLYCOCONJUGATION OF ALPHA
; FILE REFERENCE: 040853-01-5082
; CURRENT APPLICATION NUMBER: US/10/411,037
; PRIOR FILING DATE: 2003-04-09
; PRIOR APPLICATION NUMBER: US 60/328,523
; PRIOR FILING DATE: 2001-10-10
; PRIOR APPLICATION NUMBER: US 60/344,692
; PRIOR FILING DATE: 2001-10-19
; PRIOR APPLICATION NUMBER: US 60/387,292
; PRIOR FILING DATE: 2002-06-07
; PRIOR APPLICATION NUMBER: US 60/391,777
; PRIOR FILING DATE: 2002-06-25
; PRIOR APPLICATION NUMBER: US 60/396,594
; PRIOR FILING DATE: 2002-07-17
; PRIOR APPLICATION NUMBER: US 60/404,249
; PRIOR FILING DATE: 2002-08-16
; PRIOR APPLICATION NUMBER: US 60/407,527
; NUMBER OF SEQ ID NOS: 75
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 30
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-411-037-30

Query Match 99.9%; Score 12401; DB 12; Length 2351;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 2349; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 MOELSTCFFLCLAFPCFSATRYVILGAVELSWDQSDIGELPVDARPPBPVPSFPPN 60
DB 1 MOELSTCFFLCLAFPCFSATRYVILGAVELSWDQSDIGELPVDARPPBPVPSFPPN 60
QY 61 TSVVYKTLFVEFTDHLFNIAKPRPPMGLLGPFTIOAEVYDTVTWTLKNMASHPVSLHAV 120
DB 61 TSVVYKTLFVEFTDHLFNIAKPRPPMGLLGPFTIOAEVYDTVTWTLKNMASHPVSLHAV 120
QY 121 GVSYWKASEGAEYDDQTSQREKDDKVPFGSGHTYVWQVLKENGPMASDPLCLTYSYLSH 180
DB 121 GVSYWKASEGAEYDDQTSQREKDDKVPFGSGHTYVWQVLKENGPMASDPLCLTYSYLSH 180
QY 181 VDLVKDLNSGLIGALLVCREGSLAKEKTQTLHKFILLFAVDEGKSWHSETKNSLMQDRD 240
DB 181 VDLVKDLNSGLIGALLVCREGSLAKEKTQTLHKFILLFAVDEGKSWHSETKNSLMQDRD 240
QY 241 AASARAWPRGHTVNGYVARSPLGLICHRKSYVWHVIGXGTTPEVHSIFLEGHTFLVRNH 300

Db 241 AASARAPAMHVNQVIVARSIPDLGCHRKSVYVHVI GNGITPEVHSIFLEGHTLVNH 300
Qy 301 ROASLEIGPITFLTAQTILMDLQGLLSCHSSHQHDGMEAYVKVDSPCEBPOLRMKNNE 360
Db 301 ROASLEIGPITFLTAQTILMDLQGLLFCHISSHQHDGMEAYVKVDSPCEBPOLRMKNNE 360
Qy 361 BAEDYDDDLTDSMDVAFEDDONSPEFQIRSAVAKKEPKTWVHYIAAEEEDWDYAPLVLA 420
Db 361 BAEDYDDDLTDSMDVAFEDDONSPEFQIRSAVAKKEPKTWVHYIAAEEEDWDYAPLVLA 420
Qy 421 PDORSVKSOYLANGPORGRKVKVRFMAYTDETFKTRAIQHESGILGPLYLGEVGDTL 480
Db 421 PDORSVKSOYLANGPORGRKVKVRFMAYTDETFKTRAIQHESGILGPLYLGEVGDTL 480
Qy 481 LIIFKXQASRPNIYPHGTDVRLPYSRRLPKGVXHLKDFPILPGEIFKFKWTVTVEDGP 540
Db 481 LIIFKXQASRPNIYPHGTDVRLPYSRRLPKGVXHLKDFPILPGEIFKFKWTVTVEDGP 540
Qy 541 TKSDPRLCYRYSSVFVNMRDLASGLIGPLLI CYKESVDORGNOJMSDKRNILSVFDE 600
Db 541 TKSDPRLCYRYSSVFVNMRDLASGLIGPLLI CYKESVDORGNOJMSDKRNILSVFDE 600
Qy 601 NRSWILTEINIQFLPNPAGVQLEDPEFQASINMHSINGVVFDSLQLSVCLHEVAYWYILS 660
Db 601 NRSWILTEINIQFLPNPAGVQLEDPEFQASINMHSINGVVFDSLQLSVCLHEVAYWYILS 660
Qy 661 IGAQTDPLSVFSGYTFKHKMYVEDTLTLPFPGSETVFMNMPGLWILGCHNSDFNRG 720
Db 661 IGAQTDPLSVFSGYTFKHKMYVEDTLTLPFPGSETVFMNMPGLWILGCHNSDFNRG 720
Qy 721 MTALLKVSSCDKNTGVDYEDSDISAYLLSKNAJEPBSFSONSHRSTROKQFNATTI 780
Db 721 MTALLKVSSCDKNTGVDYEDSDISAYLLSKNAJEPBSFSONSHRSTROKQFNATTI 780
Qy 781 PENDLEKTDPFWAHRTPMPKIQNVSSDDLMLLRQSPTPHGLSLSDLOBAKYETFSDDPS 840
Db 781 PENDLEKTDPFWAHRTPMPKIQNVSSDDLMLLRQSPTPHGLSLSDLOBAKYETFSDDPS 840
Qy 841 PGALDNNLSLSEMTHERPQJLHSGDMVFTPESGLOJRLNEKLGTTAATBLKLLDFKVSST 900
Db 841 PGALDNNLSLSEMTHERPQJLHSGDMVFTPESGLOJRLNEKLGTTAATBLKLLDFKVSST 900
Qy 901 SNNLSTIPSDNLAACTDNTSSIGPPMPVHYDSQDITTLFGKSSPLTESGGLSLSE 960
Db 901 SNNLSTIPSDNLAACTDNTSSIGPPMPVHYDSQDITTLFGKSSPLTESGGLSLSE 960
Qy 961 NNDKLLSGLMNSQESSWGKNVSTESGRLFKGZAHGPALLTKDNALFKVSI SULKTN 1020
Db 961 NNDKLLSGLMNSQESSWGKNVSTESGRLFKGZAHGPALLTKDNALFKVSI SULKTN 1020
Qy 1021 KTSNNATNRKTHIDGFSLLIENSPPVWONILBSDTPEKKVTPLIHDMIMDKNATALRL 1080
Db 1021 KTSNNATNRKTHIDGFSLLIENSPPVWONILBSDTPEKKVTPLIHDMIMDKNATALRL 1080
Qy 1081 NMSNKTSSKNMEMVQOKKEGPIPPDAQNPDMSFFKMLFLPSARWIQRTHKNSLNSG 1140
Db 1081 NMSNKTSSKNMEMVQOKKEGPIPPDAQNPDMSFFKMLFLPSARWIQRTHKNSLNSG 1140
Qy 1141 QGSPKQVSLGPEKSVGEGNFI SEKNKVVGKGETKOVGLKEMVPPSSRNILFLTNLON 1200
Db 1141 QGSPKQVSLGPEKSVGEGNFI SEKNKVVGKGETKOVGLKEMVPPSSRNILFLTNLON 1200
Qy 1201 LHENNTNQKQKIOEELEKXETILIQENVLPQIHTVTGKNFKNLFLILSTRONVSGSYD 1260
Db 1201 LHENNTNQKQKIOEELEKXETILIQENVLPQIHTVTGKNFKNLFLILSTRONVSGSYD 1260
Qy 1261 GAYAPVLQDFRSLNDSNTRTKHTAHFSKKGEBEENLEGLOKQIVKCYACTTRISPNT 1320
Db 1261 GAYAPVLQDFRSLNDSNTRTKHTAHFSKKGEBEENLEGLOKQIVKCYACTTRISPNT 1320
Qy 1321 SQNFVTVQSRKALKQFRLPLETELEKRIIVDDTSTQWSKNMKHLTPSTLTQIDYNEKE 1380

RESULT 5

US-10-411-026-30

Db 1321 SQNFVTVQSRKALKQFRLPLETELEKRIIVDDTSTQWSKNMKHLTPSTLTQIDYNEKE 1380
Qy 1381 KGATQSPLSCLTRSHSIPQANRSPUPIAKVSSPPSRPIVLRVLFPQDNSSHLPAASY 1440
Db 1381 KGATQSPLSCLTRSHSIPQANRSPUPIAKVSSPPSRPIVLRVLFPQDNSSHLPAASY 1440
Qy 1441 RKDSGVQESSHFFQGAKKNNLSLAILTLEMTGQOREVSGSLGTSATNSVTKYKVENTVLP 1500
Db 1441 RKDSGVQESSHFFQGAKKNNLSLAILTLEMTGQOREVSGSLGTSATNSVTKYKVENTVLP 1500
Qy 1501 KPDLPKTSGKVELLPKVHIYQKDLPPPTTSNGSPGHLDLVEGSLLOQTEGAIKWNEANRP 1560
Db 1501 KPDLPKTSGKVELLPKVHIYQKDLPPPTTSNGSPGHLDLVEGSLLOQTEGAIKWNEANRP 1560
Qy 1561 GKVPFLRVATESSAKTPSKLADPLAWDNHYGTQPKBEWKSOEKSPEKTAFFKXDTILSL 1620
Db 1561 GKVPFLRVATESSAKTPSKLADPLAWDNHYGTQPKBEWKSOEKSPEKTAFFKXDTILSL 1620
Qy 1621 NACESNHAIAAINEGONKPEIEVTWAKOGRTERLCSQNPPLVKRQREITRTTLQSDQEE 1680
Db 1621 NACESNHAIAAINEGONKPEIEVTWAKOGRTERLCSQNPPLVKRQREITRTTLQSDQEE 1680
Qy 1681 IDYDDTI SVENKKSDFDIYDEDENOQSPRSFOKTRHYFIAAVERLWDYGMSSPHVLRNR 1740
Db 1681 IDYDDTI SVENKKSDFDIYDEDENOQSPRSFOKTRHYFIAAVERLWDYGMSSPHVLRNR 1740
Qy 1741 AOSGSVPQFKVVFQFETDGSFTQPLYRGELNEHLLGLLGPYIRAEVEDNIMVTFNQASR 1800
Db 1741 AOSGSVPQFKVVFQFETDGSFTQPLYRGELNEHLLGLLGPYIRAEVEDNIMVTFNQASR 1800
Qy 1801 PYSFYSSLI SYEEDQORGAEPKFNFKVNEKTYFWKQVHMAPTKDFDCXAWAYPSDV 1860
Db 1801 PYSFYSSLI SYEEDQORGAEPKFNFKVNEKTYFWKQVHMAPTKDFDCXAWAYPSDV 1860
Qy 1861 DLEKDVHSGLIGPLLVCHTNILNPAHGEQVTVQEFALFTETPDETKSYFENMERNCEA 1920
Db 1861 DLEKDVHSGLIGPLLVCHTNILNPAHGEQVTVQEFALFTETPDETKSYFENMERNCEA 1920
Qy 1921 PCNIQMEPTFKENYRFAINGYIMDTLPLVMAQDQIRWYLLSGMGNENIHSIHFSGH 1980
Db 1921 PCNIQMEPTFKENYRFAINGYIMDTLPLVMAQDQIRWYLLSGMGNENIHSIHFSGH 1980
Qy 1981 VFTVRKKEBYKMALYNLXPGVEFTEVEMLPKAGIWRVECLIGEHLHAGMSTLFLVYSNKC 2040
Db 1981 VFTVRKKEBYKMALYNLXPGVEFTEVEMLPKAGIWRVECLIGEHLHAGMSTLFLVYSNKC 2040
Qy 2041 QTPLGVASGHIRDPQITASGOYQWAPKLARLHYSGSINAWSTKEPFSWIKVDLLAPMII 2100
Db 2041 QTPLGVASGHIRDPQITASGOYQWAPKLARLHYSGSINAWSTKEPFSWIKVDLLAPMII 2100
Qy 2101 HGIKTQGAQKFSLSYISOFIIMYSLDGKKWQTYRGNSTGTILMVFPGNVDSGSIKHNI FN 2160
Db 2101 HGIKTQGAQKFSLSYISOFIIMYSLDGKKWQTYRGNSTGTILMVFPGNVDSGSIKHNI FN 2160
Qy 2161 PTIARYIRLHPHTHYSIRSTRMELMGCDLNCSPMLGMSKAI SDAQITASSYTNMFA 2220
Db 2161 PTIARYIRLHPHTHYSIRSTRMELMGCDLNCSPMLGMSKAI SDAQITASSYTNMFA 2220
Qy 2221 TWSPSKARLHLOGRSNANRPQVNNPEKMLQVDFOKTMKVTTGTQGVKSLTSMVYKBP 2280
Db 2221 TWSPSKARLHLOGRSNANRPQVNNPEKMLQVDFOKTMKVTTGTQGVKSLTSMVYKBP 2280
Qy 2281 ISSSQDGHOWTLFFQNGKVKVFCQGNQDSFTPVVNSLDPPLTRYLRIHPQSVHQAIALRM 2340
Db 2281 ISSSQDGHOWTLFFQNGKVKVFCQGNQDSFTPVVNSLDPPLTRYLRIHPQSVHQAIALRM 2340
Qy 2341 EYLGCEAODLY 2351
Db 2341 EYLGCEAODLY 2351

; Sequence 30, Application US/10411026
; Publication No. US20040063911A1
; GENERAL INFORMATION:
; APPLICANT: Neose Technologies, Inc.
; APPLICANT: DePrees, Shawn
; APPLICANT: Zopf, David
; APPLICANT: Bayer, Robert
; APPLICANT: Hakes, David
; APPLICANT: Chen, Xi
; TITLE OF INVENTION: PROTEIN REMODELING METHODS AND PROTEINS/PEPTIDES PRODUCED BY THE
; TITLE OF INVENTION: METHODS
; FILE REFERENCE: 040853-01-5053
; CURRENT APPLICATION NUMBER: US/10/411,026
; CURRENT FILING DATE: 2003-04-09
; PRIOR APPLICATION NUMBER: US 60/328,523
; PRIOR FILING DATE: 2001-10-10
; PRIOR APPLICATION NUMBER: US 60/344,692
; PRIOR FILING DATE: 2001-10-19
; PRIOR APPLICATION NUMBER: US 60/387,292
; PRIOR FILING DATE: 2002-06-07
; PRIOR APPLICATION NUMBER: US 60/391,777
; PRIOR FILING DATE: 2002-06-25
; PRIOR APPLICATION NUMBER: US 60/396,594
; PRIOR FILING DATE: 2002-07-17
; PRIOR APPLICATION NUMBER: US 60/404,249
; PRIOR FILING DATE: 2002-08-16
; PRIOR APPLICATION NUMBER: US 60/407,527
; NUMBER OF SEQ ID NOS: 75
; SOFTWARE: Patent version 3.2
; SEQ ID NO 30
; LENGTH: 2351
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-411-026-30

Query Match 99.9%; Score 12401; DB 12; Length 2351;

Best Local Similarity 99.9%; Pred. No. 0;
Matches 2349; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 MQIELSTCFELCLRFCSATRYILGAVELSWDVMQSDLGELPVDARPPRPVPSFPFN 60
DB 1 MQIELSTCFELCLRFCSATRYILGAVELSWDVMQSDLGELPVDARPPRPVPSFPFN 60
QY 61 TSVVYKTLFVEFTDHLFNIAKPRPPMGLGPTTCAEYVDTWTILKNMASHPVSLHAV 120
DB 61 TSVVYKTLFVEFTDHLFNIAKPRPPMGLGPTTCAEYVDTWTILKNMASHPVSLHAV 120
QY 121 GVSYYKASEGAYDDQTSQREKEDKVPFGSHTYVWQVLKENGPMASDPLCLTYSYLSH 180
DB 121 GVSYYKASEGAYDDQTSQREKEDKVPFGSHTYVWQVLKENGPMASDPLCLTYSYLSH 180
QY 181 VDLVKDLNSGLIGALLVCREGSLAKEKTQTLHKFILLFAVDEGKSWSETKNSLWQDRD 240
DB 181 VDLVKDLNSGLIGALLVCREGSLAKEKTQTLHKFILLFAVDEGKSWSETKNSLWQDRD 240
QY 241 AASARAWPKMTVNGYVNRSLPGLIGCRKSVYVHVIWNGTTPVHSLFEGHTFLVRNH 300
DB 241 AASARAWPKMTVNGYVNRSLPGLIGCRKSVYVHVIWNGTTPVHSLFEGHTFLVRNH 300
QY 301 ROASELISPIFTLAQTLLMDLGQFLLSCHISSHODGMEAYVKYDSCPEEPQLEMKNE 360
DB 301 ROASELISPIFTLAQTLLMDLGQFLLSCHISSHODGMEAYVKYDSCPEEPQLEMKNE 360
QY 361 EAEDYDDDLTJSEMVDVRFDDNNSPFTQIRSVAKKPKTWVHYIAABEEDWDYAPLVLA 420
DB 361 EAEDYDDDLTJSEMVDVRFDDNNSPFTQIRSVAKKPKTWVHYIAABEEDWDYAPLVLA 420
QY 421 PDRSYKSOYLNNGPQIRGRKYKVRFMAYTDETFKTREAIQHESGILGLLYGEGVDTL 480
DB 421 PDRSYKSOYLNNGPQIRGRKYKVRFMAYTDETFKTREAIQHESGILGLLYGEGVDTL 480
QY 481 LIIFKNQASRPYNIYPHGITDVRPIYSRRLPKGVXELKDFPILPGEIPFKYKWTVVEDGP 540

DB LIIFKNQASRPYNIYPHGITDVRPIYSRRLPKGVXELKDFPILPGEIPFKYKWTVVEDGP 540
QY TKS DPRCLTRYYSFVNMERDLASGLIGPLI CYKESVDQGNQIMSKRNVILFVSFDE 600
DB TKS DPRCLTRYYSFVNMERDLASGLIGPLI CYKESVDQGNQIMSKRNVILFVSFDE 600
QY NRSWYLTENIORFLPNPAGVLEDEPEFOASIMESINGYVFDLSQLSVCLHEVAYWYLS 660
DB NRSWYLTENIORFLPNPAGVLEDEPEFOASIMESINGYVFDLSQLSVCLHEVAYWYLS 660
QY IGAQTDFLSVFFSGYTPFKHVMYEDTTLTPFPFSGETVFMSENPCLMTILGCHNSDFNRNG 720
DB IGAQTDFLSVFFSGYTPFKHVMYEDTTLTPFPFSGETVFMSENPCLMTILGCHNSDFNRNG 720
QY MTALLKVSSCDKNTGDYVEDSYEDI SAYLLSKNNAIEPRSFQNSRHPSTKQKPNATTI 780
DB MTALLKVSSCDKNTGDYVEDSYEDI SAYLLSKNNAIEPRSFQNSRHPSTKQKPNATTI 780
QY PENDIEKTDPMFAHRTPMFKIQNVSSDMLLRQSPPHGILSLDLOEAKYETPSDDPS 840
DB PENDIEKTDPMFAHRTPMFKIQNVSSDMLLRQSPPHGILSLDLOEAKYETPSDDPS 840
QY PGATDSNNSISEMTHFRPQLHHSQDMVTPPSGLQLRNEKLTGTTAATLKKLDFKVSST 900
DB PGATDSNNSISEMTHFRPQLHHSQDMVTPPSGLQLRNEKLTGTTAATLKKLDFKVSST 900
QY SNNLITIPSDNLAAGTDNTSSLGPPSPMPHYDSDLTTLFGKXSSPLTSGGPLSLSEE 960
DB SNNLITIPSDNLAAGTDNTSSLGPPSPMPHYDSDLTTLFGKXSSPLTSGGPLSLSEE 960
QY NNDKLLLESGLMNSQESSWGKNSVSTBSGRLFKCKRAHGPALLTKDNALFKVSISSLKTN 1020
DB NNDKLLLESGLMNSQESSWGKNSVSTBSGRLFKCKRAHGPALLTKDNALFKVSISSLKTN 1020
QY KTSNNSATNRKTHIDGPSLLIENSFSWQNTLESSTTEPKKVTPLIHDRMLMDKNATLURL 1080
DB KTSNNSATNRKTHIDGPSLLIENSFSWQNTLESSTTEPKKVTPLIHDRMLMDKNATLURL 1080
QY NMSNKTTSKQNMWVQKKEGPIPPDAQNPDMSFFKMLFIPESARMTQRTGKNSLNSG 1140
DB NMSNKTTSKQNMWVQKKEGPIPPDAQNPDMSFFKMLFIPESARMTQRTGKNSLNSG 1140
QY CGPSFKQLVSLGPESKVEGQNFLEKKNVYVKGEBFTKDVGLKEMVFPSSNLFNLN 1200
DB CGPSFKQLVSLGPESKVEGQNFLEKKNVYVKGEBFTKDVGLKEMVFPSSNLFNLN 1200
QY LHENNTNQBKKIQEEIEKKETLIQENVVLQIHTVTGKHPMKMLFLLSTRQNVGSGYD 1260
DB LHENNTNQBKKIQEEIEKKETLIQENVVLQIHTVTGKHPMKMLFLLSTRQNVGSGYD 1260
QY LHEANNTNQBKKIQEEIEKKETLIQENVVLQIHTVTGKHPMKMLFLLSTRQNVGSGYD 1320
DB LHEANNTNQBKKIQEEIEKKETLIQENVVLQIHTVTGKHPMKMLFLLSTRQNVGSGYD 1320
QY SQONFVTOQRKALQKQRLPLEETLEKRIIVDTSTQWSXMKHLTPSTLTQIDYNEKE 1380
DB SQONFVTOQRKALQKQRLPLEETLEKRIIVDTSTQWSXMKHLTPSTLTQIDYNEKE 1380
QY KGATQSPDLSCLTRSHSIPOANRSPPLIAKVSSFPSPRPIYLRVLPODNSSHLPAASY 1440
DB KGATQSPDLSCLTRSHSIPOANRSPPLIAKVSSFPSPRPIYLRVLPODNSSHLPAASY 1440
QY RKDQSGVQESSHFIQGAKNKNSLAILLTLEMTQOREVGSGLGTGATNSVTYKKEVNTVLP 1500
DB RKDQSGVQESSHFIQGAKNKNSLAILLTLEMTQOREVGSGLGTGATNSVTYKKEVNTVLP 1500
QY KPDLPKTSIGKVELLPKVHIYQKDLFPTSTNSGSPCHLDLVEGSLQGTGEGAIKWEANRP 1560
DB KPDLPKTSIGKVELLPKVHIYQKDLFPTSTNSGSPCHLDLVEGSLQGTGEGAIKWEANRP 1560
QY GKVPFLRVAETSSAKTSPKLLDPLAWNDHYGTQPKBEWKQBSPEKTAKKKDTILSL 1620
DB GKVPFLRVAETSSAKTSPKLLDPLAWNDHYGTQPKBEWKQBSPEKTAKKKDTILSL 1620

1561 GKVPFPRVATSSAKTPSKLLDPLAWNHYGTQIPKEBWKSEKPEKTAFFKKDTILSL 1620
1621 NACBSNHAALANEGONKEEIVTWAKQRTBRLCSQNPVLRKHQBTRITRTLOSDBE 1680
1621 NACBSNHAALANEGONKEEIVTWAKQRTBRLCSQNPVLRKHQBTRITRTLOSDBE 1680
1681 IDYDDTISVEMKKEFDIYDEBENQSPRSFQKTRHYTAAVERLWDYGMSSSPHVLNR 1740
1681 IDYDDTISVEMKKEFDIYDEBENQSPRSFQKTRHYTAAVERLWDYGMSSSPHVLNR 1740
1741 AOSGVPQKVKVFOEFTDGSQTPLYRGELNEHGLGCPYIRAEVNDIMVTFNQASR 1800
1741 AOSGVPQKVKVFOEFTDGSQTPLYRGELNEHGLGCPYIRAEVNDIMVTFNQASR 1800
1801 PYSFYSSLLSYBEDORQGRAPKFNFKNETKTYFWKVOHMAPTKDBEDCKAWAFSDV 1860
1801 PYSFYSSLLSYBEDORQGRAPKFNFKNETKTYFWKVOHMAPTKDBEDCKAWAFSDV 1860
1861 DLEKDVHSLGLGLVCHTNTLPAHGRQVTVQEFALPFTTIFDETKSWYFTENMERNCA 1920
1861 DLEKDVHSLGLGLVCHTNTLPAHGRQVTVQEFALPFTTIFDETKSWYFTENMERNCA 1920
1921 PCNIOMEDPTFKENYRFRHAINCYIMDTPLGLYMAQDQRIWYLLSMGNSNENIHSHPGSH 1980
1921 PCNIOMEDPTFKENYRFRHAINCYIMDTPLGLYMAQDQRIWYLLSMGNSNENIHSHPGSH 1980
1981 VFTVKKCEYKVALNLYPGVFETVEMLPKAGIWRVECLIGEHLHAGNSTILFLVYSNKC 2040
1981 VFTVKKCEYKVALNLYPGVFETVEMLPKAGIWRVECLIGEHLHAGNSTILFLVYSNKC 2040
2041 QTPGLMAGSHIRDPOTASQYQGWAPKLARLHYSGSINAWTKPFFSWIKVDLLAPMII 2100
2041 QTPGLMAGSHIRDPOTASQYQGWAPKLARLHYSGSINAWTKPFFSWIKVDLLAPMII 2100
2101 HGKTCGABOKSSLYISOFIIMVSLDGKKQTYRGNSTGTLWPGNVDSGIXHNIFN 2160
2101 HGKTCGABOKSSLYISOFIIMVSLDGKKQTYRGNSTGTLWPGNVDSGIXHNIFN 2160
2161 PPIIARYIRLHPHYISIRSLRMELMGCDLNSCMLPGMESKAISDAQITASSYFTNMPA 2220
2161 PPIIARYIRLHPHYISIRSLRMELMGCDLNSCMLPGMESKAISDAQITASSYFTNMPA 2220
2221 TWSPSKARLHLOGRSNAPQVNNPKWLOVDQFKTMKVTVGTTQGVKSLLTSMYVKEFL 2280
2221 TWSPSKARLHLOGRSNAPQVNNPKWLOVDQFKTMKVTVGTTQGVKSLLTSMYVKEFL 2280
2281 ISSSQDGHOWTLFFQNGKVKVQGNODSETPVWNSLDPPLITRYLRHPOSWHQAIALRM 2340
2281 ISSSQDGHOWTLFFQNGKVKVQGNODSETPVWNSLDPPLITRYLRHPOSWHQAIALRM 2340
2341 EYLGCBAQDLY 2351
2341 EYLGCBAQDLY 2351

RESULT 6

US-10-360-101-229
; Sequence 229, Application US/10360101
; Publication No. US20040009550A1
; GENERAL INFORMATION:
; APPLICANT: Moll, Gert N.
; APPLICANT: Leenhouts, Cornelis J.
; TITLE OF INVENTION: Export and modification of (poly)peptide in the lantibiotic way
; FILE REFERENCE: 2183-5673
; CURRENT APPLICATION NUMBER: US/10/360,101
; CURRENT FILING DATE: 2003-02-07
; PRIOR APPLICATION NUMBER: EP 02077060.8
; PRIOR FILING DATE: 2002-05-24
; NUMBER OF SEQ ID NOS: 309
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 229
; LENGTH: 2332
; TYP3: PRT

; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: sequence of factor VIII
US-10-360-101-229
Query Match 99.1%; Score 12306; DB 15; Length 2332;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2331; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 20 ATRRYVLCVELSNDYQSDIGELPVDARFPVRPKSPFNSTSVYKKTLVEFTDHLFN 79
DB 1 ATRRYVLCVELSNDYQSDIGELPVDARFPVRPKSPFNSTSVYKKTLVEFTDHLFN 60
QY 80 IAKRPPPMWGLLGPITQAEVYDVTVITLKNMASHPVSLHAGVSVKASGAEDDQTSQ 139
DB 61 IAKRPPPMWGLLGPITQAEVYDVTVITLKNMASHPVSLHAGVSVKASGAEDDQTSQ 120
QY 140 REKDDKVPFGGSETYVWQVLKENGPMASDPLCTYSYLSHVDLVKDLNSGLIGALLVCR 199
DB 121 REKDDKVPFGGSETYVWQVLKENGPMASDPLCTYSYLSHVDLVKDLNSGLIGALLVCR 180
QY 200 EGSIAKEKTOTLHKPILLFAVDEGKSWHSETKSLMODRDAASARAKPKHETVNGYNR 259
DB 181 EGSIAKEKTOTLHKPILLFAVDEGKSWHSETKSLMODRDAASARAKPKHETVNGYNR 240
QY 260 SLPLGLIGCHRXSVYWHVIGMTTPEVHSIFLEGHTFLVNRHQASLEISPIFTTAQTLL 319
DB 241 SLPLGLIGCHRXSVYWHVIGMTTPEVHSIFLEGHTFLVNRHQASLEISPIFTTAQTLL 300
QY 320 MDLGGFLLSCHISHQHDGMEAYVKVDCPBEPPQIRKKNHEAEYDDDLTDSMDVVRP 379
DB 301 MDLGGFLFCHISHQHDGMEAYVKVDCPBEPPQIRKKNHEAEYDDDLTDSMDVVRP 360
QY 380 DDNSPSFIQIRSAKHPKTHVYIAAEEDWDYAPLVAPDDRSYKSOYLNGPQIRG 439
DB 361 DDNSPSFIQIRSAKHPKTHVYIAAEEDWDYAPLVAPDDRSYKSOYLNGPQIRG 420
QY 440 RYKXVRFMAYTDETFKTRAIQHESGILGELLXGEVGTLLIIFKNQASRPYNIYPGI 499
DB 421 RYKXVRFMAYTDETFKTRAIQHESGILGELLXGEVGTLLIIFKNQASRPYNIYPGI 480
QY 500 TDVRLYGRRLPKGVKHUKDPFLLPGEIFKXKVTWVDEGDTKSDPRCLTRYSSFYVME 559
DB 481 TDVRLYGRRLPKGVKHUKDPFLLPGEIFKXKVTWVDEGDTKSDPRCLTRYSSFYVME 540
QY 560 RDLASGLIGPLLICVKESVDORGNQIMSKDNVILFSVDENRSMYLTENQRFLPNPAG 619
DB 541 RDLASGLIGPLLICVKESVDORGNQIMSKDNVILFSVDENRSMYLTENQRFLPNPAG 600
QY 620 VOLEDPPEFQASNMHSINGYVFDLSQLSVCLHEVAYWYILSIGAQTDFLSVFFSGYTFKH 679
DB 601 VOLEDPPEFQASNMHSINGYVFDLSQLSVCLHEVAYWYILSIGAQTDFLSVFFSGYTFKH 660
QY 680 XWYVEDTTLTPFSGETVMSXENPGLMILGCHNSDPENRGMTALLKVSSCDKNTGDYVE 739
DB 661 XWYVEDTTLTPFSGETVMSXENPGLMILGCHNSDPENRGMTALLKVSSCDKNTGDYVE 720
QY 740 DSYEDISAYLLSKNNAIEPRSPSONSRHPSFTRQKQFNATTPENDIEKTDPMFAHRTMP 799
DB 721 DSYEDISAYLLSKNNAIEPRSPSONSRHPSFTRQKQFNATTPENDIEKTDPMFAHRTMP 780
QY 800 KIQNVSSDDLMLLRQSTPPLGLSLDIOEAKYETFSDDPSGAIDSNNSISEMTHFRPQ 859
DB 781 KIQNVSSDDLMLLRQSTPPLGLSLDIOEAKYETFSDDPSGAIDSNNSISEMTHFRPQ 840
QY 860 LHSQGMVFTESGQLRLNEKLGTAAATLKLDPKVSSTNNLITIPSDNLAAGTDN 919
DB 841 LHSQGMVFTESGQLRLNEKLGTAAATLKLDPKVSSTNNLITIPSDNLAAGTDN 900
QY 920 TSSLGPPGMPVHYDSQLODTTLFGKKSPLTBSGGPLSSENNDSKLLSGLKNSQESSW 979
DB 901 TSSLGPPGMPVHYDSQLODTTLFGKKSPLTBSGGPLSSENNDSKLLSGLKNSQESSW 960

QY 980 GKNVSTESGRFLKGRAGHGPALLTKDNALFKVSIISLLKTNKTSNNASATNRKTHIDGPSL 1039
Db 961 GKNVSTESGRFLKGRAGHGPALLTKONALFKVSIISLLKTNKTSNNASATNRKTHIDGPSL 1020
QY 1040 LEENSPSVWQNLILESDTEPKKVTPLIHDRMLMDKATALRLNMGKNTTSSKMMWVQOK 1099
Db 1021 LEENSPSVWQNLILESDTEPKKVTPLIHDRMLMDKATALRLNMGKNTTSSKMMWVQOK 1080
QY 1100 KEGPIPPAQNDPMGFFKMLFLPESARWIORTHGKNSLNSGGSPKQOLVSI-GPKSVEG 1159
Db 1081 KEGPIPPAQNDPMGFFKMLFLPESARWIORTHGKNSLNSGGSPKQOLVSI-GPKSVEG 1140
QY 1160 QNFLSEKKNVVGKGEFTKDVGLKEMVPPSSRNLFITLNDLNHNNTNQEKKI0EEIEK 1219
Db 1141 QNFLSEKKNVVGKGEFTKDVGLKEMVPPSSRNLFITLNDLNHNNTNQEKKI0EEIEK 1200
QY 1220 KETLLQENNVLPQIHTVGTGNFMKMLLSTRQNVGSGYDAPVLODFRSLNDSNR 1279
Db 1201 KETLLQENNVLPQIHTVGTGNFMKMLLSTRQNVGSGYDAPVLODFRSLNDSNR 1260
QY 1280 TKKHTAHFSKKEEENLEGLGNQTKQIWEKACTTRISPTNSQONFVFORSGKALKQREL 1339
Db 1261 TKKHTAHFSKKEEENLEGLGNQTKQIWEKACTTRISPTNSQONFVFORSGKALKQREL 1320
QY 1340 PLEETELEKRIIVDDTSTOWSKMKHLPSTLTQIDYNEKKGKAITQGPLSDCLTRSHSI 1399
Db 1321 PLEETELEKRIIVDDTSTOWSKMKHLPSTLTQIDYNEKKGKAITQGPLSDCLTRSHSI 1380
QY 1400 POANRSPPIAKVSFPRIPIYITRVLFDNSSHLPAASTRKDSGVQOESSHFLQGAKK 1459
Db 1381 POANRSPPIAKVSFPRIPIYITRVLFDNSSHLPAASTRKDSGVQOESSHFLQGAKK 1440
QY 1460 NNLISAILTLLEMIGQDRVGSGLTSATNSVYKVENLTKPDLPKTSKGVELLPKVHI 1519
Db 1441 NNLISAILTLLEMIGQDRVGSGLTSATNSVYKVENLTKPDLPKTSKGVELLPKVHI 1500
QY 1520 YOKDLFPETNSGPHLDLVEGSLLOQTEGALKWNEANRPCKVPLFRVATESAKTFSK 1579
Db 1501 YOKDLFPETNSGPHLDLVEGSLLOQTEGALKWNEANRPCKVPLFRVATESAKTFSK 1560
QY 1580 LLDPLANDNHVGTQIPKEWKSQSKPEKTAFAKKTITLSNACSNHAIAINEGQONKP 1639
Db 1561 LLDPLANDNHVGTQIPKEWKSQSKPEKTAFAKKTITLSNACSNHAIAINEGQONKP 1620
QY 1640 EIEVTWAKQGRTERLCSQNPPLVRKHOREIIRTTLSQDOEBIDYDDTISVEMKXEDFDIY 1699
Db 1621 EIEVTWAKQGRTERLCSQNPPLVRKHOREIIRTTLSQDOEBIDYDDTISVEMKXEDFDIY 1680
QY 1700 DEDENQSPRSFQKTRHYFTIAAVERLDYDGMSSSPHVLNRNQAQSGVPPQFKVVFQBFDT 1759
Db 1681 DEDENQSPRSFQKTRHYFTIAAVERLDYDGMSSSPHVLNRNQAQSGVPPQFKVVFQBFDT 1740
QY 1760 GSFTQPLRYGSLNHLGLGPYIEAEVEDNIMVTFRQAQSPYFYSGLISYEDQOQGA 1819
Db 1741 GSFTQPLRYGSLNHLGLGPYIEAEVEDNIMVTFRQAQSPYFYSGLISYEDQOQGA 1800
QY 1820 EPRKNFVKPNETKTYFWKQVQHMAPTKDEDFCKAWAYFSDVLEKDVHSGLIGPLLVCHT 1879
Db 1801 EPRKNFVKPNETKTYFWKQVQHMAPTKDEDFCKAWAYFSDVLEKDVHSGLIGPLLVCHT 1860
QY 1880 NTLNPAHGRQVTOEFALFTTIPDETQSWTPTENMERNCRAPCNQIOMEDTFFKENVYFHA 1939
Db 1861 NTLNPAHGRQVTOEFALFTTIPDETQSWTPTENMERNCRAPCNQIOMEDTFFKENVYFHA 1920
QY 1940 INGVIMDTPLCLVMAQOQRIWYLLSGNSNENIHSIHFSGHVFTVRKKEEYKXALNLYP 1999
Db 1921 INGVIMDTPLCLVMAQOQRIWYLLSGNSNENIHSIHFSGHVFTVRKKEEYKXALNLYP 1980
QY 2000 GVFETVEMLPKAGIWRVECLIGEBLHAGMSTLFLVYSNKCQTPGLMGASHINDFOQTAS 2059
Db 1981 GVFETVEMLPKAGIWRVECLIGEBLHAGMSTLFLVYSNKCQTPGLMGASHINDFOQTAS 2040
QY 2060 GOYGQWAPKLARLHYSGSINAWSTKEPFSWKVDLLAPMIHGIKTOGARQKFSLSYISQ 2119

Db 2041 GOYGQWAPKLARLHYSGSINAWSTKEPFSWKVDLLAPMIHGIKTOGARQKFSLSYISQ 2100
QY 2120 FIIMYSLDGKKWQYRGNGSTGTMLVFFGNVDSSGIGKHNIFNPPIIARVIRLHPTHSIRS 2179
Db 2101 FIIMYSLDGKKWQYRGNGSTGTMLVFFGNVDSSGIGKHNIFNPPIIARVIRLHPTHSIRS 2160
QY 2180 TLRBELMGCDLNSCSMPLGMSKAISSDAQITASSYFTNMPATWSPSKARLHLQGRSNWR 2239
Db 2161 TLRBELMGCDLNSCSMPLGMSKAISSDAQITASSYFTNMPATWSPSKARLHLQGRSNWR 2220
QY 2240 PQVNRPKEWLQVDFOKTMKVTVGTQGVKSLLTSMYKKEFLISSODGHOWTLFQNGKV 2299
Db 2221 PQVNRPKEWLQVDFOKTMKVTVGTQGVKSLLTSMYKKEFLISSODGHOWTLFQNGKV 2280
QY 2300 KVFQGNQDSFTPVVNSLDPPLLTRYLRHQPQSVHQAIALRMEVLGCEAQDLY 2351
Db 2281 KVFQGNQDSFTPVVNSLDPPLLTRYLRHQPQSVHQAIALRMEVLGCEAQDLY 2332
RESULT 7
US-09-957-641-2
; Sequence 2, Application US/09957641
; Publication No. US20020182670A1
; GENERAL INFORMATION:
; APPLICANT: Emory University
; TITLE OF INVENTION: MODIFIED FACTOR VIII
; FILE REFERENCE: 75-00
; CURRENT APPLICATION NUMBER: US/09/957,641
; CURRENT FILING DATE: 2001-09-16
; PRIOR APPLICATION NUMBER: US 60/234047
; PRIOR FILING DATE: 2000-09-19
; PRIOR APPLICATION NUMBER: US 60/236460
; PRIOR FILING DATE: 2000-09-29
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 2332
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-957-641-2
Query Match 99.0%; Score 12293; DB 9; Length 2332;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 2329; Conservativa 1; Mismatches 2; Indels 0; Gaps 0;
QY 20 ATRRYLGAVELGWDYMQSDLGELPVDARPPRPVPSFPFNTSVVYKTLFVETDHLFN 79
Db 1 ATRRYLGAVELGWDYMQSDLGELPVDARPPRPVPSFPFNTSVVYKTLFVETDHLFN 60
QY 80 IAKPRPPMGLGPTIOAEVYDVTVTILKNVASHPVSLHAGVSYWKASGAEYDDQTSQ 139
Db 61 IAKPRPPMGLGPTIOAEVYDVTVTILKNVASHPVSLHAGVSYWKASGAEYDDQTSQ 120
QY 140 REKEDDKVPFGSGHTYVWQVLKENGPMASDPLCLTYSYLSHVDLVKDLNSGLIGALLVCR 199
Db 121 REKEDDKVPFGSGHTYVWQVLKENGPMASDPLCLTYSYLSHVDLVKDLNSGLIGALLVCR 180
QY 200 EGSIAKEKTOHLKPIILLFAVFDGKSWHSETKNSLMQDRDAASARAWPKMHTVNGVNR 259
Db 181 EGSIAKEKTOHLKPIILLFAVFDGKSWHSETKNSLMQDRDAASARAWPKMHTVNGVNR 240
QY 260 SLPLGLGCHKSVYHVGITGTPVHSIFLEGTFLVRNHRQASLEISPIITLTAQTLL 319
Db 241 SLPLGLGCHKSVYHVGITGTPVHSIFLEGTFLVRNHRQASLEISPIITLTAQTLL 300
QY 320 MDLGQFLLSHSSHQDGMVAVYKVDSCPEBQPMKNNEAEYDDDLTDSMDVVR 379
Db 301 MDLGQFLLSHSSHQDGMVAVYKVDSCPEBQPMKNNEAEYDDDLTDSMDVVR 360
QY 380 DDNSPSFTQIRSAVAKKPKTWHYIAAEEEDWDYAPLVLPDDRYSKQYLNNGPQIRG 439
Db 361 DDNSPSFTQIRSAVAKKPKTWHYIAAEEEDWDYAPLVLPDDRYSKQYLNNGPQIRG 420

QY 440 RYKVVREMYATDETEKREAIQHESGILGPIYGEVGDITLIIIPKNOASRPYNIPYHGI 499
 DB 421 RYKVVREMYATDETEKREAIQHESGILGPIYGEVGDITLIIIPKNOASRPYNIPYHGI 480
 QY 500 TDVRLYRRRLPKGVKHLKDPILPGEIFKYKWTVTVDGPTKSDPRCLTRYYSFVNE 559
 DB 481 TDVRLYRRRLPKGVKHLKDPILPGEIFKYKWTVTVDGPTKSDPRCLTRYYSFVNE 540
 QY 560 RDLASGLIGPILICYKESVDQGNQIMSDKRVILFVFDENRSWYLNTQRFNPAG 619
 DB 541 RDLASGLIGPILICYKESVDQGNQIMSDKRVILFVFDENRSWYLNTQRFNPAG 600
 QY 620 VOLEDEPEFOASIMHSINGYVFDISQLSVCLHEVAYWYILSCAQTDFLSVFFSGYTKH 679
 DB 601 VOLEDEPEFOASIMHSINGYVFDISQLSVCLHEVAYWYILSCAQTDFLSVFFSGYTKH 660
 QY 680 KMVYEDTTLTFPFGSETVFMSENPGILNIGCHNSDFNRGMNTALLKVSSCDKNTGDYIE 739
 DB 661 KMVYEDTTLTFPFGSETVFMSENPGILNIGCHNSDFNRGMNTALLKVSSCDKNTGDYIE 720
 QY 740 DSYEDI SAVILLSKNAIIPRSFSQNSRHPSTKQFNATTIPENDIEKTDWFAHRTMP 799
 DB 721 DSYEDI SAVILLSKNAIIPRSFSQNSRHPSTKQFNATTIPENDIEKTDWFAHRTMP 780
 QY 800 KIQNVSSDILLMLRQSPHPHGLSLSDLOEAKYETFSDDPSGAIDSNNSLSEMTHF3PQ 859
 DB 781 KIQNVSSDILLMLRQSPHPHGLSLSDLOEAKYETFSDDPSGAIDSNNSLSEMTHF3PQ 840
 QY 860 LHSQDMVFTPSGLOLRNEXKLTGTAATELKLDFKVSSTNNLITSTPSNLAAGTDN 919
 DB 841 LHSQDMVFTPSGLOLRNEXKLTGTAATELKLDFKVSSTNNLITSTPSNLAAGTDN 900
 QY 920 TSSLGPPSMVHYDOLDTTLFGKSSPSTESGGPLSLSEENNDKLLBSGLMNSQESSW 979
 DB 901 TSSLGPPSMVHYDOLDTTLFGKSSPSTESGGPLSLSEENNDKLLBSGLMNSQESSW 960
 QY 980 GKNVSTESGRIFKGRAGHPALLTKDNALFKVSISSLKTNKTSNNKTNKTHIDGPSL 1039
 DB 961 GKNVSTESGRIFKGRAGHPALLTKDNALFKVSISSLKTNKTSNNKTNKTHIDGPSL 1020
 QY 1040 LIENSPVWONILESTESKKTPTPLIHDRMLDKNATLRLHMSKNTTSSKNMVMQOK 1099
 DB 1021 LIENSPVWONILESTESKKTPTPLIHDRMLDKNATLRLHMSKNTTSSKNMVMQOK 1080
 QY 1100 KEGPDPDAQNDMFPFKMLFLPESARWIQRTGKNSLNSGOGPPKQVLSLGPESKSVBG 1159
 DB 1081 KEGPDPDAQNDMFPFKMLFLPESARWIQRTGKNSLNSGOGPPKQVLSLGPESKSVBG 1140
 QY 1160 QNPLSEKNKVYVGKGFRTKDVGLKENVFPSSRNLFITNLDNLHENNTHNQEKKIQEETEK 1219
 DB 1141 QNPLSEKNKVYVGKGFRTKDVGLKENVFPSSRNLFITNLDNLHENNTHNQEKKIQEETEK 1200
 QY 1220 KETLLOENVVLPQIHVTCTKFMKNLFLSTRQNVGSGYDAGAYAPVLOPFSLNDSTNR 1279
 DB 1201 KETLLOENVVLPQIHVTCTKFMKNLFLSTRQNVGSGYDAGAYAPVLOPFSLNDSTNR 1260
 QY 1280 TKKHTAHFSGKGEENLEGLGNQTKQIVKAYACTTRISNTSQONFVQORSKRALKQFRL 1339
 DB 1261 TKKHTAHFSGKGEENLEGLGNQTKQIVKAYACTTRISNTSQONFVQORSKRALKQFRL 1320
 QY 1340 PLEETELEKRIIVDDTSTQWSKMKHLTPSTLTQIDYNEKEKGAITQSPSLCLTRSHSI 1399
 DB 1321 PLEETELEKRIIVDDTSTQWSKMKHLTPSTLTQIDYNEKEKGAITQSPSLCLTRSHSI 1380
 QY 1400 PQANRSPLEPIAKVSSPSPSRPIYLTVLQDNSSHLPAASYYKKGDSGVQESSHFLOGAKK 1459
 DB 1381 PQANRSPLEPIAKVSSPSPSRPIYLTVLQDNSSHLPAASYYKKGDSGVQESSHFLOGAKK 1440
 QY 1460 NNLSLAILTLEMTGDOREVGSLGTSATNSVYKVENTVLPKPDLPKTSKVELLPKVHI 1519
 DB 1441 NNLSLAILTLEMTGDOREVGSLGTSATNSVYKVENTVLPKPDLPKTSKVELLPKVHI 1500

QY 1520 YQKDLFTPTSTNGSPGHILDLVEGSLLOCTEGAIKXNEANREPKVPFLRVATESAKTPSK 1579
 DB 1501 YQKDLFTPTSTNGSPGHILDLVEGSLLOCTEGAIKXNEANREPKVPFLRVATESAKTPSK 1560
 QY 1580 LLDPLANDNHVGTQIPKEEWKSQESKPAKTKDITILSNACESHAIAMINEGQNK 1639
 DB 1561 LLDPLANDNHVGTQIPKEEWKSQESKPAKTKDITILSNACESHAIAMINEGQNK 1620
 QY 1640 BIEVTWAKQGRTERLCSNPPVLKXHOREIRTRTLOSQOBEIDYDDTILSVEMKEDDPDIY 1699
 DB 1621 BIEVTWAKQGRTERLCSNPPVLKXHOREIRTRTLOSQOBEIDYDDTILSVEMKEDDPDIY 1680
 QY 1700 DEDENQSPRSQKTRHYFIAAVERLNDYGMSSPHVLNRNAQSGSVFPQKVVVFQETD 1759
 DB 1681 DEDENQSPRSQKTRHYFIAAVERLNDYGMSSPHVLNRNAQSGSVFPQKVVVFQETD 1740
 QY 1760 CSFTQPLYRGELNEHLGILGPYIRAEVEDNTMVFRNAQASRPYFSSLSIYEEDQROGA 1819
 DB 1741 CSFTQPLYRGELNEHLGILGPYIRAEVEDNTMVFRNAQASRPYFSSLSIYEEDQROGA 1800
 QY 1820 EPRKXVKNPNTKTYFWKVQVHMAPTKDFDCKAWAYESDVKDHSVGLIGPLLVCHT 1879
 DB 1801 EPRKXVKNPNTKTYFWKVQVHMAPTKDFDCKAWAYESDVKDHSVGLIGPLLVCHT 1860
 QY 1880 NTLNPAHGRQVTVGEFALLFTIFDITKSWYFTENMERNCRAPCNQIMEDPTFKENYRFEA 1939
 DB 1861 NTLNPAHGRQVTVGEFALLFTIFDITKSWYFTENMERNCRAPCNQIMEDPTFKENYRFEA 1920
 QY 1940 INGYIMDTLPGVNAQODORIWYLLSMGSENHISIHESGHVFTVRKKEEYKMALYNLYP 1999
 DB 1921 INGYIMDTLPGVNAQODORIWYLLSMGSENHISIHESGHVFTVRKKEEYKMALYNLYP 1980
 QY 2000 GYFTEVEMLPKAGIWRVECLIGEHLHAGMGTFLVYSNKQCTPLGWSAGEIRDPQITAS 2059
 DB 1981 GYFTEVEMLPKAGIWRVECLIGEHLHAGMGTFLVYSNKQCTPLGWSAGEIRDPQITAS 2040
 QY 2060 QYQGWAPKLAARLHYSGSINAWSTKEPFSWIKVOLLAPMI IHGKTQGAROKFSSLYISQ 2119
 DB 2041 QYQGWAPKLAARLHYSGSINAWSTKEPFSWIKVOLLAPMI IHGKTQGAROKFSSLYISQ 2100
 QY 2120 FIIMYSLDGKKWOTYRGNSTGTLMVFFGNVDSSGKKNHFNPPHARVIRLHPHYSIRS 2179
 DB 2101 FIIMYSLDGKKWOTYRGNSTGTLMVFFGNVDSSGKKNHFNPPHARVIRLHPHYSIRS 2160
 QY 2180 TIRMEIMGCDLNSCSPYGLMESKAISSDAQITASSYFTNKFATWSPSKARLHLQGRSNAR 2239
 DB 2161 TIRMEIMGCDLNSCSPYGLMESKAISSDAQITASSYFTNKFATWSPSKARLHLQGRSNAR 2220
 QY 2240 POWNPKEWLQVDFQKTKMKTGVTTQGVKSLITSYVKEFLISSQDGHQWTLFPQNGKV 2299
 DB 2221 POWNPKEWLQVDFQKTKMKTGVTTQGVKSLITSYVKEFLISSQDGHQWTLFPQNGKV 2280
 QY 2300 KVFQGNQDSFTPVVNSLDPPLTLRLTHPOSVWHQIALRMEVLGCEAODLY 2351
 DB 2281 KVFQGNQDSFTPVVNSLDPPLTLRLTHPOSVWHQIALRMEVLGCEAODLY 2332

RESULT 8

US-10-187-319-2
 ; Sequence 2, Application US/10187319
 ; Publication NO. US20030068785A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Lolilar, John S.
 ; TITLE OF INVENTION: Hybrid Human/Animal Factor VIII
 ; NUMBER OF SEQUENCES: 40
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSSEE: Greenlee, Winner and Sullivan, P.C.
 ; STREET: 5370 Manhattan Circle Suite 201
 ; CITY: Boulder
 ; STATE: Colorado
 ; COUNTRY: USA
 ; ZIP: 80303
 ; COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/10/187,319
FILING DATE: 27-Aug-2002
CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 09/523,656
FILING DATE: 2000-03-10
APPLICATION NUMBER: US 09/037,601
FILING DATE: 1998-03-10
APPLICATION NUMBER: WO PCT/US97/11155
FILING DATE: 1997-06-26
APPLICATION NUMBER: US 08/670,707
FILING DATE: 1996-06-26

ATTORNEY/AGENT INFORMATION:

NAME: Greenlee, Lorraine L.
REGISTRATION NUMBER: 27,894
REFERENCE/DOCKET NUMBER: 75-95K
TELECOMMUNICATION INFORMATION:
TELEPHONE: 303/499-8080
TELEFAX: 303/499-8089

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:
LENGTH: 2332 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: <unknown>
MOLECULE TYPE: protein
HYPOTHETICAL: YES
ANTI-SENSE: NO
FRAGMENT TYPE: N-terminal
ORIGINAL SOURCE:

ORGANISM: Homo sapiens

TISSUE TYPE: Liver

SEQUENCE DESCRIPTION: SEQ ID NO: 2:

US-10-187-319-2

Query Match 99.0%; Score 12293; DB 14; Length 2332;

Best Local Similarity 99.9%; Pred. No. 0;

Matches 2329; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 20 ATRRYVIGAVLSWDMYSDIGELPVDARFPRVPKSPBNTSVYKKTILFVEFTDHLFN 79
DB 1 ATRRYVIGAVLSWDMYSDIGELPVDARFPRVPKSPBNTSVYKKTILFVEFTDHLFN 60
QY 80 IAKRPPPMGLGPTIOAEVYDVTITLKNMASHFVSLHAGVSVYKASEGAEYDDQTSQ 139
DB 61 IAKRPPPMGLGPTIOAEVYDVTITLKNMASHFVSLHAGVSVYKASEGAEYDDQTSQ 120
QY 140 REKEDDKVPGGSHYVWQVLKENGPMASDPLCLTYSLSHVDLVKDLNSGLIGALLVCR 199
DB 121 REKEDDKVPGGSHYVWQVLKENGPMASDPLCLTYSLSHVDLVKDLNSGLIGALLVCR 180
QY 200 EGSIAKEXTOTLHKFTILLFAVFDGKSWHSETKNSLMODRDAASARAWPKHVTNGVYNR 259
DB 181 EGSIAKEXTOTLHKFTILLFAVFDGKSWHSETKNSLMODRDAASARAWPKHVTNGVYNR 240
QY 260 SLPLGLIGHRSKSVYVHVTGMGTTPEVHGI FLEGHTFLVNRHQASLEISPTFLTAQTL 319
DB 241 SLPLGLIGHRSKSVYVHVTGMGTTPEVHGI FLEGHTFLVNRHQASLEISPTFLTAQTL 300
QY 320 MDLQGLFLSHCHISSHOHDGMEAYKVDCSPEPQLRMKQNEAEADYDDDLTDSMDVVRP 379
DB 301 MDLQGLFLSHCHISSHOHDGMEAYKVDCSPEPQLRMKQNEAEADYDDDLTDSMDVVRP 360
QY 380 DDNDSPSPQIRSVAKKPKTWVHYIAAEEEDWDYAPLVLPADORSYKSOYLNNGPQRI 439
DB 361 DDNDSPSPQIRSVAKKPKTWVHYIAAEEEDWDYAPLVLPADORSYKSOYLNNGPQRI 420
QY 440 RYKVKVRMAVTDTEFKTREAIQHSGLPLLYGEVGDTHLIIIFKNQASRPYNIYPHGI 499

DB 421 RYKVKVRMAVTDTEFKTREAIQHSGLPLLYGEVGDTHLIIIFKNQASRPYNIYPHGI 480
QY 500 TDVRPLYSRRLPKGVKHLKDFPILPGEIFKYKWTVTVEDGPTKSDPRCLTRYSSFVAME 559
DB 481 TDVRPLYSRRLPKGVKHLKDFPILPGEIFKYKWTVTVEDGPTKSDPRCLTRYSSFVAME 540
QY 560 RDLASGLIGPLLI CYKESVDORGNOIMDKKNVILFVSFVDENRSWYLTENTQRFLEPAG 619
DB 541 RDLASGLIGPLLI CYKESVDORGNOIMDKKNVILFVSFVDENRSWYLTENTQRFLEPAG 600
QY 620 VQLEDPEFQASINMHSINGYVFDLSQLSVCLHEVAYWVILSIGAOTDLSVFFESYTKH 679
DB 601 VQLEDPEFQASINMHSINGYVFDLSQLSVCLHEVAYWVILSIGAOTDLSVFFESYTKH 660
QY 680 KXVVEDTLTLFPFSGETVFMSENPGMWILQCHNSDFENRGMTALLKVSSCDKNTGYE 739
DB 661 KXVVEDTLTLFPFSGETVFMSENPGMWILQCHNSDFENRGMTALLKVSSCDKNTGYE 720
QY 740 DSYEDISAYILSKNNAIEPRSPQNSRHPSTROQFNATTIPENDIEKTDWFAHRTMP 799
DB 721 DSYEDISAYILSKNNAIEPRSPQNSRHPSTROQFNATTIPENDIEKTDWFAHRTMP 780
QY 800 KIQNVSSDILLMLLRQSPTHGLSLSDLOEAKYETFSDDPSGALDSNNSLSEMTHRPQ 859
DB 781 KIQNVSSDILLMLLRQSPTHGLSLSDLOEAKYETFSDDPSGALDSNNSLSEMTHRPQ 840
QY 860 LHHSGDMVFTPESGLOLRINEKLGTTATELKLDFKYSSSTNNLIPTIPSDNLAAGTDN 919
DB 841 LHHSGDMVFTPESGLOLRINEKLGTTATELKLDFKYSSSTNNLIPTIPSDNLAAGTDN 900
QY 920 TSSLGPPSPMPVHYDSQDITLFGKSSDLTSGGPLSSENNDSKLLSEGLMNSQESSW 979
DB 901 TSSLGPPSPMPVHYDSQDITLFGKSSDLTSGGPLSSENNDSKLLSEGLMNSQESSW 960
QY 980 GKNVSTSTESGRLFKGKRAHGPAALLTKONALFKVSIILKTKNTKNSATNREKTHIDGPSL 1039
DB 961 GKNVSTSTESGRLFKGKRAHGPAALLTKONALFKVSIILKTKNTKNSATNREKTHIDGPSL 1020
QY 1040 LIENSPSVQWQNILSDTEFKKVTPLIHDRMLMDKNATRLNHNMSKNTTSSKNMVMQOK 1099
DB 1021 LIENSPSVQWQNILSDTEFKKVTPLIHDRMLMDKNATRLNHNMSKNTTSSKNMVMQOK 1080
QY 1100 KEGPIPDQAQNPDMSPFKMLFLPESARWI QRTGKNSLNSCQGPSKQVSLGPEKSV 1159
DB 1081 KEGPIPDQAQNPDMSPFKMLFLPESARWI QRTGKNSLNSCQGPSKQVSLGPEKSV 1140
QY 1160 QNFLSEKNKVVVGKGEFTKDVGLKEMVFPSSRNFLTNLNLNHNNTNNOEKKIOEIEK 1219
DB 1141 QNFLSEKNKVVVGKGEFTKDVGLKEMVFPSSRNFLTNLNLNHNNTNNOEKKIOEIEK 1200
QY 1220 KETLIQENVVLPOHTVTGTQKFMKNJFLLSSTRQNVSGSYDGAAPVLQDPRSLNDSNR 1279
DB 1201 KETLIQENVVLPOHTVTGTQKFMKNJFLLSSTRQNVSGSYDGAAPVLQDPRSLNDSNR 1260
QY 1280 TKKHTAFPSKGEENLEGLGNQTKQVEKACATRI SPNTSQONFVTOFSKRAKQFRL 1339
DB 1261 TKKHTAFPSKGEENLEGLGNQTKQVEKACATRI SPNTSQONFVTOFSKRAKQFRL 1320
QY 1340 PLEETELEKRII VDDTSTQWSKNMKHLPSTLTQIDYNEKEKGAI TQSPSDCLTRSHSI 1399
DB 1321 PLEETELEKRII VDDTSTQWSKNMKHLPSTLTQIDYNEKEKGAI TQSPSDCLTRSHSI 1380
QY 1400 PQANRSLPIAKVSSPFIPIYITRVLFQDNSSHLPAASVYRKDQSGVQSSSHFLQAKX 1459
DB 1381 PQANRSLPIAKVSSPFIPIYITRVLFQDNSSHLPAASVYRKDQSGVQSSSHFLQAKX 1440
QY 1460 NNLSLAII LTEMTCQOREVSLGTSATNSVTKKVENTVLPKDLPKTSGKVELLPKVI 1519
DB 1441 NNLSLAII LTEMTCQOREVSLGTSATNSVTKKVENTVLPKDLPKTSGKVELLPKVI 1500
QY 1520 YQKDLFTETSNGPSGHLDLVEGSLLOCTBGAIKWNEANRPGKVPFLRVATESSAKTPSK 1579

1501 YQKDLFPETSGNSGPHGLDLVEGSLLOCTEGAIKWEANRRPKGVFLRVATATESAKTPSK 1560
1580 LLDPLANDHHVGTQIPKSEWKSQKSPKTAFAKKDDTILSNACSNHAIINAGQNK 1639
1561 LLDPLANDHHVGTQIPKSEWKSQKSPKTAFAKKDDTILSNACSNHAIINAGQNK 1620
1640 BIEVTWAGQRTERCSQNPVVKHQRREIRITLQSQDEEDYDDTISVENKKEDFDIY 1699
1621 BIEVTWAGQRTERCSQNPVVKHQRREIRITLQSQDEEDYDDTISVENKKEDFDIY 1680
1700 DEDENCSPRSFQKTRHVFIAAVERLWYGMSSSPHYLRNRAQSGSVPOFKKVPQETD 1759
1681 DEDENCSPRSFQKTRHVFIAAVERLWYGMSSSPHYLRNRAQSGSVPOFKKVPQETD 1740
1760 GSFTQPLVGEINELHGLLGPVIRAEVDENIMVTFRNOASRPSYSSILSVEEDQROGA 1819
1741 GSFTQPLVGEINELHGLLGPVIRAEVDENIMVTFRNOASRPSYSSILSVEEDQROGA 1800
1820 EPRKNFVKENETKTYFWKVQHNAPTKDEFFCKAWAYFSVDVLEKDVHSGLIGPLLVCHT 1879
1601 EPRKNFVKENETKTYFWKVQHNAPTKDEFFCKAWAYFSVDVLEKDVHSGLIGPLLVCHT 1860
1880 NTLNPAHGRQVTVQBFALFFFTIFDETKSWYFTENMERNCRAPCNQMEDPTFKENYRPHA 1939
1861 NTLNPAHGRQVTVQBFALFFFTIFDETKSWYFTENMERNCRAPCNQMEDPTFKENYRPHA 1920
1940 INGYIMDTLPGLVMAQDQIRWYLLSMGSENIHSHFSGHVFYTKKBYKVALYNYLP 1999
1921 INGYIMDTLPGLVMAQDQIRWYLLSMGSENIHSHFSGHVFYTKKBYKVALYNYLP 1980
2000 GYFVTEMLPSKAGWRYVECLIGEHLHAGMSTFLVYVSNKQCTPLGMASGHIRDFQITAS 2059
1981 GYFVTEMLPSKAGWRYVECLIGEHLHAGMSTFLVYVSNKQCTPLGMASGHIRDFQITAS 2040
2060 GYGQWAPKARLHVSNGSNASTKEPFSWIKVDLLAPMIHGKTOGARQKPSLIYSQ 2119
2041 GYGQWAPKARLHVSNGSNASTKEPFSWIKVDLLAPMIHGKTOGARQKPSLIYSQ 2100
2120 FTMYSLDGKKQWTRGNSGTGLVYFFGNVDSGKHNIFNPPIIARYIRLHPTHYSIRS 2179
2101 FTMYSLDGKKQWTRGNSGTGLVYFFGNVDSGKHNIFNPPIIARYIRLHPTHYSIRS 2160
2180 TLRMELMGCDLNSCNSMPLGMSKASIDAOITASSYFTNNFATWSPSKALHLQGRSNWR 2239
2161 TLRMELMGCDLNSCNSMPLGMSKASIDAOITASSYFTNNFATWSPSKALHLQGRSNWR 2220
2240 POWNPKENLOYDFOKTMAVTVGTQGVKSLLTSMVYKBEFLISSODGHQWTLFQNGKV 2299
2221 POWNPKENLOYDFOKTMAVTVGTQGVKSLLTSMVYKBEFLISSODGHQWTLFQNGKV 2280
2300 KVFQGNQDSFTPVNSLDPLFLRYLRHFPQSVWHQIALRMEVLCGEAQDLY 2351
2281 KVFQGNQDSFTPVNSLDPLFLRYLRHFPQSVWHQIALRMEVLCGEAQDLY 2332

RESULT 9
US-10-131-510A-2
; Sequence 2, Application US/10131510A
; Publication No. US20030166536A1
; GENERAL INFORMATION:
; APPLICANT: Lollar, John S
; TITLE OF INVENTION: Modified Factor VIII
; FILE REFERENCE: 75-95J
; CURRENT APPLICATION NUMBER: US/10/131,510A
; CURRENT FILING DATE: 2002-07-10
; PRIOR APPLICATION NUMBER: U.S. 09/315,179
; PRIOR FILING DATE: 1999-05-20
; PRIOR APPLICATION NUMBER: U.S. 09/037,601
; PRIOR FILING DATE: 1998-03-10
; PRIOR APPLICATION NUMBER: U.S. 08/670,707
; PRIOR FILING DATE: 1996-06-26
; PRIOR APPLICATION NUMBER: PCT/US97/11155
; PRIOR FILING DATE: 1997-06-26

PRIOR APPLICATION NUMBER: PCT/US94/13200
; PRIOR FILING DATE: 1994-11-15
; PRIOR APPLICATION NUMBER: U.S. 08/212,133
; PRIOR FILING DATE: 1994-03-11
; PRIOR APPLICATION NUMBER: U.S. 07/864,004
; PRIOR FILING DATE: 1992-04-07
; NUMBER OF SEQ ID NOS: 40
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 2332
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-10-131-510A-2

Query Match 99.0%; Score 12293; DB 14; Length 2332;
Best local similarity 99.9%; Ered. No. 0;
Matches 2329; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 20 ATRRYYLGAVELSDWYQSDIGELPVDARPPRPVKGPFPNTSVYKTLFVEFTDHLFN 79
DB 1 ATRRYYLGAVELSDWYQSDIGELPVDARPPRPVKGPFPNTSVYKTLFVEFTDHLFN 60

QY 80 IAKRPPPMGLLGTIOAEVYDVTNITLKNASHPVSLHAGVSYWKASEGAYDDQTSQ 139
DB 61 IAKRPPPMGLLGTIOAEVYDVTNITLKNASHPVSLHAGVSYWKASEGAYDDQTSQ 120

QY 140 REKEDDKVFPGSGSHTYVMQVLKENGPMASDPLCLTYSYLSHVDLVKOLNSGLIGALLVCR 199
DB 121 REKEDDKVFPGSGSHTYVMQVLKENGPMASDPLCLTYSYLSHVDLVKOLNSGLIGALLVCR 180

QY 200 EGSIAKEKTOQLHFKPILLPAVFDGKSWHSTKSLMODRDAASARAWPKMTVNGYNR 259
DB 181 EGSIAKEKTOQLHFKPILLPAVFDGKSWHSTKSLMODRDAASARAWPKMTVNGYNR 240

QY 260 SLPLGLIGCHRSVYVWVGMGTTPBHVHSIFLEGHTFLVNRHQASLEISPIITLTAQTL 319
DB 241 SLPLGLIGCHRSVYVWVGMGTTPBHVHSIFLEGHTFLVNRHQASLEISPIITLTAQTL 300

QY 320 MDLGOFLLSCHISSHQHDMBAYVKVDCSPEPQLRMKNNBEAEDYDDLDSEMDVYRF 379
DB 301 MDLGOFLLSCHISSHQHDMBAYVKVDCSPEPQLRMKNNBEAEDYDDLDSEMDVYRF 360

QY 380 DDNSPSPTQIRSNVAKGPKTWVYIAAEBEDWDYAPVLAPDDRSYSKSYLNNGPQIRG 439
DB 361 DDNSPSPTQIRSNVAKGPKTWVYIAAEBEDWDYAPVLAPDDRSYSKSYLNNGPQIRG 420

QY 440 RYKXKVRFMAYTDETFKTRAIQHSGLGPLLYGEVGTLLIFKNQASRPYNIYPHGI 499
DB 421 RYKXKVRFMAYTDETFKTRAIQHSGLGPLLYGEVGTLLIFKNQASRPYNIYPHGI 480

QY 500 TDVRLYRRRLPKGVKHLKDPFIPLPGEIFKYKWTVTVEDGPTKSDPRCLTRYSSFVAME 559
DB 481 TDVRLYRRRLPKGVKHLKDPFIPLPGEIFKYKWTVTVEDGPTKSDPRCLTRYSSFVAME 540

QY 560 RDLASGLIGPLLIICYKESVDQGNQIMSKDNVILFSVPDENRSKYLNTORFLPNPAG 619
DB 541 RDLASGLIGPLLIICYKESVDQGNQIMSKDNVILFSVPDENRSKYLNTORFLPNPAG 600

QY 620 VQLEDPEFQASNIHMSINGYVDFSQLSVCLHEVAYWILSICAGTDFLSVFFSGYTPKH 679
DB 601 VQLEDPEFQASNIHMSINGYVDFSQLSVCLHEVAYWILSICAGTDFLSVFFSGYTPKH 660

QY 680 KMVYEDTLTLFPFSGETVPMNMENPGLWILGCHNSDFNRNGMTALLKVSQCDKNTGDIYE 739
DB 661 KMVYEDTLTLFPFSGETVPMNMENPGLWILGCHNSDFNRNGMTALLKVSQCDKNTGDIYE 720

QY 740 DSYEDISAYLLSKNNAIEPRSFQSRHPSQKQFNATITPENDIEKTDPAHRTMP 799
DB 721 DSYEDISAYLLSKNNAIEPRSFQSRHPSQKQFNATITPENDIEKTDPAHRTMP 780

QY 800 KIQNVSSSDLLMLRQSPHGLSLSDIQEAKYETFSDDPSGAIDSNNSLSEMTFRPQ 859
DB 781 KIQNVSSSDLLMLRQSPHGLSLSDIQEAKYETFSDDPSGAIDSNNSLSEMTFRPQ 840

QY 860 LHHSGDMVFPESGLQRLNEKLGTTAATELKKLDKVSSTNNLSTIPSDNLAAGTDN 919
DB 841 LHHSGDMVFPESGLQRLNEKLGTTAATELKKLDKVSSTNNLSTIPSDNLAAGTDN 900
QY 920 TSSLGPPMPVHYDQSDTLTGKKSGLTBSGGPLSGSEENNDKLESGLMNSQSSW 979
DB 901 TSSLGPPMPVHYDQSDTLTGKKSGLTBSGGPLSGSEENNDKLESGLMNSQSSW 960
QY 980 GKNVSSTESGLFGKKAHGPALLTKONALFKVSI SLLKTNKTNNATRKTHIDGPEL 1039
DB 961 GKNVSSTESGLFGKKAHGPALLTKONALFKVSI SLLKTNKTNNATRKTHIDGPEL 1020
QY 1040 LIENSPPVWQNI LESDTEFKVTPLIHDMMDKONATRLNEMSNKTTSSKNMEMVQOK 1099
DB 1021 LIENSPPVWQNI LESDTEFKVTPLIHDMMDKONATRLNEMSNKTTSSKNMEMVQOK 1080
QY 1100 KEGIPDDAQNPMDSFFROMLFLPSARWIOQTHGKNSLNSQGPSKQLVSLGPEKSV 1159
DB 1081 KEGIPDDAQNPMDSFFROMLFLPSARWIOQTHGKNSLNSQGPSKQLVSLGPEKSV 1140
QY 1160 QNFUSEKNKVVGKGETKOVGLKEMVFPSSRNFLTNLDNLHNNTHNQEKKEQBI 1219
DB 1141 QNFUSEKNKVVGKGETKOVGLKEMVFPSSRNFLTNLDNLHNNTHNQEKKEQBI 1200
QY 1220 KETLIQENVVLPOHTVTGTRKNFMKNFLASTRONVSGSYDGAAPVLQDPRSLN 1279
DB 1201 KETLIQENVVLPOHTVTGTRKNFMKNFLASTRONVSGSYDGAAPVLQDPRSLN 1260
QY 1280 TKKHTAHFSKGBEENLEGLNQTKQIVKVIACATRTISPNTSQONFVQSKRALKQFRL 1339
DB 1261 TKKHTAHFSKGBEENLEGLNQTKQIVKVIACATRTISPNTSQONFVQSKRALKQFRL 1320
QY 1340 PLEETELEKRIIVDDTSTQSKNKHITPSTLTQIDYNEKXKGAITQSPISDCITRSHSI 1399
DB 1321 PLEETELEKRIIVDDTSTQSKNKHITPSTLTQIDYNEKXKGAITQSPISDCITRSHSI 1380
QY 1400 POANRSLPIAKVSSPFSIRPIYLRVLFDONSSHLPAASRYKKDQSGVQSSSHFLOGAKK 1459
DB 1381 POANRSLPIAKVSSPFSIRPIYLRVLFDONSSHLPAASRYKKDQSGVQSSSHFLOGAKK 1440
QY 1460 NNLSLAITLTEMTCQREVGLSTSAINSVTYKVENIVLPKPDLPKTSKVLELLPKVHI 1519
DB 1441 NNLSLAITLTEMTCQREVGLSTSAINSVTYKVENIVLPKPDLPKTSKVLELLPKVHI 1500
QY 1520 YQKDLFFPTEISNGSPGHLDLVEGSLLOCTEGAIKXNEANRPGKVPFLRVATESSAKTPSK 1579
DB 1501 YQKDLFFPTEISNGSPGHLDLVEGSLLOCTEGAIKXNEANRPGKVPFLRVATESSAKTPSK 1560
QY 1580 LLDPLAWDNHYGTQIPKEBWKQSKSPEKTAPEKCKDTILSLNACESNHAIAAINEGQNK 1639
DB 1561 LLDPLAWDNHYGTQIPKEBWKQSKSPEKTAPEKCKDTILSLNACESNHAIAAINEGQNK 1620
QY 1640 EIEVTWAKQSTELCSQNPVLKSHOREITRTTLOSQDEIDYDDPISVEMKEDFDIY 1699
DB 1621 EIEVTWAKQSTELCSQNPVLKSHOREITRTTLOSQDEIDYDDPISVEMKEDFDIY 1680
QY 1700 DEDENQSPRSFQKTRHYFIAAVERLMDYGMSSSPHYLRNRAQSGVPQPKVVFQBF 1759
DB 1681 DEDENQSPRSFQKTRHYFIAAVERLMDYGMSSSPHYLRNRAQSGVPQPKVVFQBF 1740
QY 1760 GSTQPIYRGELNEHGLGOPYTRAENVEDNIMVTRNQAQSRPYSFYSSLLSYBEDQOQA 1819
DB 1741 GSTQPIYRGELNEHGLGOPYTRAENVEDNIMVTRNQAQSRPYSFYSSLLSYBEDQOQA 1800
QY 1820 EPRKNFVKPNETKTYFWKVOHHMAPTDBDFCKAWAYFSVDVLEKOVHSGLIGELLVCHT 1879
DB 1801 EPRKNFVKPNETKTYFWKVOHHMAPTDBDFCKAWAYFSVDVLEKOVHSGLIGELLVCHT 1860
QY 1880 NTLNPAHGRQVTVQEFALFTTIFDETCKSWYFTENWERNCRAPCNQIOMEDPTFKENYRFA 1939
DB 1861 NTLNPAHGRQVTVQEFALFTTIFDETCKSWYFTENWERNCRAPCNQIOMEDPTFKENYRFA 1920

QY 1940 INGVIYMDTLPEGLYMAQDQRIRWYLLSMGSENTHSIHPSGHVFTVRKKEEYKMALYNLYP 1999
DB 1921 INGVIYMDTLPEGLYMAQDQRIRWYLLSMGSENTHSIHPSGHVFTVRKKEEYKMALYNLYP 1980
QY 2000 GVFETVEMLPKAGIWRVECLIGEHJHAGMSTLFLVYSNKCQTPGLGVASGHIRDFOITAS 2059
DB 1981 GVFETVEMLPKAGIWRVECLIGEHJHAGMSTLFLVYSNKCQTPGLGVASGHIRDFOITAS 2040
QY 2060 GOYGOWAPKLARLHYSGSINAWSTKEPFSWKVDLLAPMIHGIKTOGAKOKFSSLYISQ 2119
DB 2041 GOYGOWAPKLARLHYSGSINAWSTKEPFSWKVDLLAPMIHGIKTOGAKOKFSSLYISQ 2100
QY 2120 FIWYSLDGKKWOTYRGNSTGTLWPFNGVDSGIGIKENIENPPIIARYIRLHPHYIRS 2179
DB 2101 FIWYSLDGKKWOTYRGNSTGTLWPFNGVDSGIGIKENIENPPIIARYIRLHPHYIRS 2160
QY 2180 TLRMELMGCDLNSCSPMLGVESKAISDAQITASSYFTNMFATWSPSKARLHLOGRSNAR 2239
DB 2161 TLRMELMGCDLNSCSPMLGVESKAISDAQITASSYFTNMFATWSPSKARLHLOGRSNAR 2220
QY 2240 POVNNPKEMLOVDPOKTMKVTVGTTOGVKSLLTSMYKFEFLISSQDGHQWTLFFQNGKV 2299
DB 2221 POVNNPKEMLOVDPOKTMKVTVGTTOGVKSLLTSMYKFEFLISSQDGHQWTLFFQNGKV 2280
QY 2300 KVFQGNODSTPVPVNSLDPELLTRYLRHPQSWHQAIALRMEVLGCBAQDLY 2351
DB 2281 KVFQGNODSTPVPVNSLDPELLTRYLRHPQSWHQAIALRMEVLGCBAQDLY 2332

RESULT 10

US-10-445-235-2
; Sequence 2, Application: US/10445235
; Publication No. US2004000570A1
; GENERAL INFORMATION:
; APPLICANT: Katherine A. High
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE
; FILE REFERENCE: CHOP0176
; CURRENT APPLICATION NUMBER: US/10/445,235
; PRIOR FILING DATE: 2003-05-22
; PRIOR FILING DATE: 2002-05-22
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: Fast-Seq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 2332
; TYPE: PRT
; ORGANISM: homo sapiens
US-10-445-235-2

Query Match 99.0%; Score 12293; DB 15; Length 2332;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 2329; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 20 ATRRYVLGAVELSWDMQSDLGELPVDARPPVPKSPFPNTGVYKTKLFEVFTDHLFN 79
DB 1 ATRRYVLGAVELSWDMQSDLGELPVDARPPVPKSPFPNTGVYKTKLFEVFTDHLFN 60
QY 80 IAKPRPPMGLLQPTTQAEVYDVTWITLKNMASHPVSLHAGVSYWKASGAEYDDQTSQ 139
DB 61 IAKPRPPMGLLQPTTQAEVYDVTWITLKNMASHPVSLHAGVSYWKASGAEYDDQTSQ 120
QY 140 REKEDDKVPPGGSHTYVWQVLKENGSPWSDPLCLTYSYLSHVDLVKDLNSGLIGALLVCR 199
DB 121 REKEDDKVPPGGSHTYVWQVLKENGSPWSDPLCLTYSYLSHVDLVKDLNSGLIGALLVCR 180
QY 200 EGSIAKEKQTTLHKFTILLFAVFDGKSNHSETKNSLMQDRDAASARAWPKQHTWNGVNR 259
DB 181 EGSIAKEKQTTLHKFTILLFAVFDGKSNHSETKNSLMQDRDAASARAWPKQHTWNGVNR 240
QY 260 SIPGLTGCURKSVYHVHVGNGTTPPEVHSIFLEGHTFLVRNHRQASLEISPIITLTAQTLL 319

Db 241 SLPLGLGCHRKSVYWHVIGMTTPBHVHSIFLEGHTFLVRNHRQASLEISPIITFLTAQTLL 300
Qy 320 MDLGGFLLSCHISSHQHDMGMEAYKVYVDSQBPBQLRMKNREAEYDDDLTDSMDVYRF 379
Db 301 MDLGGFLLSCHISSHQHDMGMEAYKVYVDSQBPBQLRMKNREAEYDDDLTDSMDVYRF 360
Qy 380 DDNSPFSFIQRSVAKGPKTWVHIAAAEEEDWDYAPLAPDDRSYSQYLNNGPQIG 439
Db 361 DDNSPFSFIQRSVAKGPKTWVHIAAAEEEDWDYAPLAPDDRSYSQYLNNGPQIG 420
Qy 440 RYKXVRFMAYTDEFTKTRAEIOHESGILGPLYGEVGDTLIIIFKNOASRPYNYPHGI 499
Db 421 RYKXVRFMAYTDEFTKTRAEIOHESGILGPLYGEVGDTLIIIFKNOASRPYNYPHGI 480
Qy 500 TDVRLYSRRLPKGVKHLKDFILPGEIFPKYKWTYVBDGPTKSDPRCLTRYSSFVAME 559
Db 481 TDVRLYSRRLPKGVKHLKDFILPGEIFPKYKWTYVBDGPTKSDPRCLTRYSSFVAME 540
Qy 560 RDLASGLIGPLIICYKESVDQRGNQIMSDKRNVLFSVPDENRWSYLTENIORPLNDAG 619
Db 541 RDLASGLIGPLIICYKESVDQRGNQIMSDKRNVLFSVPDENRWSYLTENIORPLNDAG 600
Qy 620 VOLEDPEFOASNMHSINGYVFDLSQLSVCLHEVAYWYILSTGAOTDFLSVFFSGYTFKH 679
Db 601 VOLEDPEFOASNMHSINGYVFDLSQLSVCLHEVAYWYILSTGAOTDFLSVFFSGYTFKH 660
Qy 680 KMVYEDTLTLFPFSGETVMSMENPCLWILGCHNSDFRNRGNTALLKYSSCDKNTGDYIE 739
Db 661 KMVYEDTLTLFPFSGETVMSMENPCLWILGCHNSDFRNRGNTALLKYSSCDKNTGDYIE 720
Qy 740 DSYEDISAYLLSKNNAIEPRSPSONSRHESTRQKQFNATTIPENDIEKTDPMFAHRTMP 799
Db 721 DSYEDISAYLLSKNNAIEPRSPSONSRHESTRQKQFNATTIPENDIEKTDPMFAHRTMP 780
Qy 800 KIQNVSSDLLMLLKQSPTHGSLSDLOEAKYETFSDDPSGALDSNNLSLSEMTHPRPQ 859
Db 781 KIQNVSSDLLMLLKQSPTHGSLSDLOEAKYETFSDDPSGALDSNNLSLSEMTHPRPQ 840
Qy 860 LHSQDMVTPSSGLQRLNEXLGTAAETELKLPFKVSTSNLIISTIPSNLAAGTDN 919
Db 841 LHSQDMVTPSSGLQRLNEXLGTAAETELKLPFKVSTSNLIISTIPSNLAAGTDN 900
Qy 920 TSSLGPPSPMEVHYDQDQTLTLFGKXSSPLTESGGPLSLSEENNDKLLBSGLMNSQESSW 979
Db 901 TSSLGPPSPMEVHYDQDQTLTLFGKXSSPLTESGGPLSLSEENNDKLLBSGLMNSQESSW 960
Qy 980 GKNVSTESGRILPKGRAGPALLTKONALFKVSIISLTKNTSNNSATNRKTHIDGPSL 1039
Db 961 GKNVSTESGRILPKGRAGPALLTKONALFKVSIISLTKNTSNNSATNRKTHIDGPSL 1020
Qy 1040 LIENSPSVWQNIILESTEFKKTVPILHDMRLMDKATALLRNMKNKTTSSKNMKNVQOK 1099
Db 1021 LIENSPSVWQNIILESTEFKKTVPILHDMRLMDKATALLRNMKNKTTSSKNMKNVQOK 1080
Qy 1100 KEGPIPPDAQNDPMGFFKMLFPEGARWIORTHGKNSLNSGQGPQKQVLSLGPESKVEG 1159
Db 1081 KEGPIPPDAQNDPMGFFKMLFPEGARWIORTHGKNSLNSGQGPQKQVLSLGPESKVEG 1140
Qy 1160 QNPLSEKNVWVKGEBTQDVGLKENVFPSSRNLPFTNLIDNHNHNTHQEKIOEIEBK 1219
Db 1141 QNPLSEKNVWVKGEBTQDVGLKENVFPSSRNLPFTNLIDNHNHNTHQEKIOEIEBK 1200
Qy 1220 KETLIQENVVLPOIHTVGTQKFMKNLFLILSTRQNVESYDQAYAPVLQDFSLNDSINR 1279
Db 1201 KETLIQENVVLPOIHTVGTQKFMKNLFLILSTRQNVESYDQAYAPVLQDFSLNDSINR 1260
Qy 1280 TKKHTAHFKGKEENLEGLNQTKOIVBKYACTTRISPNTSQQNFVTOQRSKRLKQFPL 1339
Db 1261 TKKHTAHFKGKEENLEGLNQTKOIVBKYACTTRISPNTSQQNFVTOQRSKRLKQFPL 1320
Qy 1340 PLEETELEKRIIIVDDTSTOWSKNMKHLTFTSTLTDIDYNEKEKGAITQSPSLCLTRSHSI 1399
Db 1321 PLEETELEKRIIIVDDTSTOWSKNMKHLTFTSTLTDIDYNEKEKGAITQSPSLCLTRSHSI 1380

Qy 1400 PQANRSLPIAKVSSPFSIRPIYLTIVLFDNSSHLPAASVYRKXDSGVQBSHSLQGAKK 1459
Db 1381 PQANRSLPIAKVSSPFSIRPIYLTIVLFDNSSHLPAASVYRKXDSGVQBSHSLQGAKK 1440
Qy 1460 NNLSLAILTLEMTGDQREVGSIGTSAINSVTYKVENTVLPKPDLPKTSKGVKVELLPKVI 1519
Db 1441 NNLSLAILTLEMTGDQREVGSIGTSAINSVTYKVENTVLPKPDLPKTSKGVKVELLPKVI 1500
Qy 1520 YQKDLFFTETNSGPGHLDLVEGSLLOSTEGAIKWNANRPCKVPFFLRVATESAKTPSK 1579
Db 1501 YQKDLFFTETNSGPGHLDLVEGSLLOSTEGAIKWNANRPCKVPFFLRVATESAKTPSK 1560
Qy 1580 LLDPLANDNHVGTQIPKEEWSQESKPEKTAFAKKDITLSLNACSHHAIINAGQNK 1639
Db 1561 LLDPLANDNHVGTQIPKEEWSQESKPEKTAFAKKDITLSLNACSHHAIINAGQNK 1620
Qy 1640 EIEVTWAKQGRTERLCSQNPPLVKRHOEITRTTLQSDQEEIDYDDTISVEMKKEDEDIY 1699
Db 1621 EIEVTWAKQGRTERLCSQNPPLVKRHOEITRTTLQSDQEEIDYDDTISVEMKKEDEDIY 1680
Qy 1700 DEDNQSPRSQKTRHYFIAAVERLWDYGVSSPHVLNRBAQSGSVQPKKVPQBTBD 1759
Db 1681 DEDNQSPRSQKTRHYFIAAVERLWDYGVSSPHVLNRBAQSGSVQPKKVPQBTBD 1740
Qy 1760 GSFTQPLVYRGELNHLGLLGPIYIAEVEDNIMVTFRNOASRPYFYSSLSIYEBDQOGA 1819
Db 1741 GSFTQPLVYRGELNHLGLLGPIYIAEVEDNIMVTFRNOASRPYFYSSLSIYEBDQOGA 1800
Qy 1820 EPRXNFVKPNETKTYFWKVQHMAPTKDEFCCKAWAYFSDVDLEKDVHSGHIGLLVCHT 1879
Db 1801 EPRXNFVKPNETKTYFWKVQHMAPTKDEFCCKAWAYFSDVDLEKDVHSGHIGLLVCHT 1860
Qy 1880 NTLPAHGRQVTVGEFALPFTIFDETSKWYFTEENMERNCRAPONTOMEDPTKENYRPHA 1939
Db 1861 NTLPAHGRQVTVGEFALPFTIFDETSKWYFTEENMERNCRAPONTOMEDPTKENYRPHA 1920
Qy 1940 INGYIMDTPLGLVWAQDQIRWYLLSMGSENHIIHPSGHVFTVRKKEZYKMALYNLYP 1999
Db 1921 INGYIMDTPLGLVWAQDQIRWYLLSMGSENHIIHPSGHVFTVRKKEZYKMALYNLYP 1980
Qy 2000 GVFETVEMLPKAGIWRVECLIGEHLHAGMTFLVYSNKKCQOTPLGMASGHIRDPQITAS 2059
Db 1981 GVFETVEMLPKAGIWRVECLIGEHLHAGMTFLVYSNKKCQOTPLGMASGHIRDPQITAS 2040
Qy 2060 QOYGOWAPKLARLHYSGSINAWSTKEPESWIKVDLAPMIHIGIKTOGAROKFSSLYISQ 2119
Db 2041 QOYGOWAPKLARLHYSGSINAWSTKEPESWIKVDLAPMIHIGIKTOGAROKFSSLYISQ 2100
Qy 2120 FIIMYSLDGKXQVYRGNSGTGLVFFGNVDSSGKHNIPNPIIARIYIRLHPHYSIRS 2179
Db 2101 FIIMYSLDGKXQVYRGNSGTGLVFFGNVDSSGKHNIPNPIIARIYIRLHPHYSIRS 2160
Qy 2180 TLRLMELMGCDLNSCMLPGMESKAIISDAQITASSYFTNMFATWSPSKARLHQRSNAWR 2239
Db 2161 TLRLMELMGCDLNSCMLPGMESKAIISDAQITASSYFTNMFATWSPSKARLHQRSNAWR 2220
Qy 2240 PQVNNPKMLQVDFQKTKWKTGVTTOGVKSLLTSMYVKEPLISSODGHOWTLFFQNGKV 2299
Db 2221 PQVNNPKMLQVDFQKTKWKTGVTTOGVKSLLTSMYVKEPLISSODGHOWTLFFQNGKV 2280
Qy 2300 KVFQGNQDSFPPVNSLDPPLLTLYRIRHPQSVWHQIALRMEVLGCEAQDLY 2351
Db 2281 KVFQGNQDSFPPVNSLDPPLLTLYRIRHPQSVWHQIALRMEVLGCEAQDLY 2332

RESULT 11
US-10-239-498A-2
; Sequence 2, Application US/10239498A
; Publication No. US2004002333A1
; GENERAL INFORMATION:
; APPLICANT: Hauser, Charlotte
; APPLICANT: Horster, Andrea

APPLICANT: Schroder, Carol
APPLICANT: Lehner, Michael
TITLE OF INVENTION: Production of Recombinant Blood Clotting Factors in
TITLE OF INVENTION: Human Cell Lines
FILE REFERENCE: 80977.0001
CURRENT APPLICATION NUMBER: US/10/239,498A
CURRENT FILING DATE: 2003-07-08
PRIOR APPLICATION NUMBER: PCT/EP01/03220
NUMBER OF SEQ ID NOS: 21
SOFTWARE: Patent In Ver. 2.1
SEQ ID NO 2
LENGTH: 2332
TYPE: PRT
ORGANISM: Homo sapiens
US-10-239-498A-2

Query Match 99.0%; Score 12293; DB 16; Length 2332;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 2329; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 20 ATRRYLGAVALSWDMQSDLGELPVDARPPRPVPSRPENTSVVYKTLFVEFTHLNP 79
DB 1 ATRRYLGAVALSWDMQSDLGELPVDARPPRPVPSRPENTSVVYKTLFVEFTHLNP 60

QY 80 TAPRPPWMLLGPITQAEVYTVITLKNMASHPVSLHAAGVSVYKASGAEYDDQTSQ 139
DB 61 TAPRPPWMLLGPITQAEVYTVITLKNMASHPVSLHAAGVSVYKASGAEYDDQTSQ 120

QY 140 REKEDKVPFGGSHYVVOVLKENGPMASDPCLTYSLYSHVDLVKDLNSGLIGALLVCR 199
DB 121 REKEDKVPFGGSHYVVOVLKENGPMASDPCLTYSLYSHVDLVKDLNSGLIGALLVCR 180

QY 200 EGSIAKKTOTLHKITLLFAVDEGKSWSEKNSLMQDRDAASAPAWPKMTVNGYVR 259
DB 181 EGSIAKKTOTLHKITLLFAVDEGKSWSEKNSLMQDRDAASAPAWPKMTVNGYVR 240

QY 260 SLPGILGCHRKSVYVHVIGWGTTPVHSIFLEGHTFLVNRHQASLEISPIITLTAQTLL 319
DB 241 SLPGILGCHRKSVYVHVIGWGTTPVHSIFLEGHTFLVNRHQASLEISPIITLTAQTLL 300

QY 320 MDLQGLLSCHTSSHQHDMBAIVKVDSCPEPQLKYNKNEAEADYDDDLTISEMDVVRP 379
DB 301 MDLQGLLFCHTSSHQHDMBAIVKVDSCPEPQLKYNKNEAEADYDDDLTISEMDVVRP 360

QY 360 DDNSPSFIQIRSVAKHPEKTWWHYIAAEEEDWDYAPLVLAEDDSYKSCYLLNNGPQIRG 439
DB 361 DDNSPSFIQIRSVAKHPEKTWWHYIAAEEEDWDYAPLVLAEDDSYKSCYLLNNGPQIRG 420

QY 440 RYKRVRFMAYTDETPKTEATQCHESGILGPLLYGEVGTLLIIIFKQASRPNTYPHGI 499
DB 421 RYKRVRFMAYTDETPKTEATQCHESGILGPLLYGEVGTLLIIIFKQASRPNTYPHGI 480

QY 500 TDVRLYSRRLPKGVKHLXDFPLPCEIFKYKWTVTVEGPKSDPRCLTRYYSFVNME 559
DB 481 TDVRLYSRRLPKGVKHLXDFPLPCEIFKYKWTVTVEGPKSDPRCLTRYYSFVNME 540

QY 560 RDLASGLIGPLLIICYKESVDQGNQIMSKRNVLFSVFDENRSWYLTENIQFPLNPAG 619
DB 541 RDLASGLIGPLLIICYKESVDQGNQIMSKRNVLFSVFDENRSWYLTENIQFPLNPAG 600

QY 620 VQLEDEPQASINMSINGVYFDSQLSVCLHEVAWYILSTGAQDFFLSVFFSGYTPKH 679
DB 601 VQLEDEPQASINMSINGVYFDSQLSVCLHEVAWYILSTGAQDFFLSVFFSGYTPKH 660

QY 680 KMVYEDTLILFPFSGTVMENPGWLILGCHNSDFRNGMTALLKVSCKNKTGYYE 739
DB 661 KMVYEDTLILFPFSGTVMENPGWLILGCHNSDFRNGMTALLKVSCKNKTGYYE 720

QY 740 DSYEDISAYLLSKNAIEPRSPQNSRHPSTROKQFNATTIPENDIEKTDPMFAHRTMP 799
DB 721 DSYEDISAYLLSKNAIEPRSPQNSRHPSTROKQFNATTIPENDIEKTDPMFAHRTMP 780

QY 800 KIQVSSDILLMLLRQSTPHGLSLDLOEAKYETFSDDPGALDSNNLSSEMTFRPQ 859
DB 781 KIQVSSDILLMLLRQSTPHGLSLDLOEAKYETFSDDPGALDSNNLSSEMTFRPQ 840

QY 860 LHSQDMVFTPESGLQLRLNEKLGTTAATELKKLDPKVSSSTNNLIISTIPDNLAAGTDN 929
DB 841 LHSQDMVFTPESGLQLRLNEKLGTTAATELKKLDPKVSSSTNNLIISTIPDNLAAGTDN 900

QY 920 TSSLGPPSPVHYDSQDQDTTLFGKSSPLTSSGGPLSLSEENNDKSLLESGLMNSQESSW 979
DB 901 TSSLGPPSPVHYDSQDQDTTLFGKSSPLTSSGGPLSLSEENNDKSLLESGLMNSQESSW 960

QY 980 GKNVSTESGRLFKGKRAHGPALLTKONALFKVSI-SLLKTKNTKNSNATNKTHTIDGPSL 1039
DB 961 GKNVSTESGRLFKGKRAHGPALLTKONALFKVSI-SLLKTKNTKNSNATNKTHTIDGPSL 1020

QY 1040 LIENSPTSVMQILSDTBFKVTPLIHDRMLDMKNATALRLNHNKNTSSKNMVMQOK 1099
DB 1021 LIENSPTSVMQILSDTBFKVTPLIHDRMLDMKNATALRLNHNKNTSSKNMVMQOK 1080

QY 1100 KEGPIPPDAQPDMSFFKMLFLPESARMIQTHGKNSLNSGGPSPKOLVSLGPESKVEG 1159
DB 1081 KEGPIPPDAQPDMSFFKMLFLPESARMIQTHGKNSLNSGGPSPKOLVSLGPESKVEG 1140

QY 1160 QNFLSEKKNVVVGKEFTKDVGLKEMVFPSSRNFLTNLDMLNHNTHNQEKKIQEEIEK 1219
DB 1141 QNFLSEKKNVVVGKEFTKDVGLKEMVFPSSRNFLTNLDMLNHNTHNQEKKIQEEIEK 1200

QY 1220 KETLLQENNVLPQHTVTGKTNFMKNLFLSTRQNVESYDGAAYAPVLQDPRSLNDSNR 1279
DB 1201 KETLLQENNVLPQHTVTGKTNFMKNLFLSTRQNVESYDGAAYAPVLQDPRSLNDSNR 1260

QY 1280 TKKTHAHSKXGEBENLEGLGNOTKQIVKACTTRISPNTSQQNFVTRQSRKALKQREL 1339
DB 1261 TKKTHAHSKXGEBENLEGLGNOTKQIVKACTTRISPNTSQQNFVTRQSRKALKQREL 1320

QY 1340 PLEETEELKRIIVDDTSTQWSKNMKHLTPSTLTQIDYNEKEKGAITQSPSLDCLTRSHI 1399
DB 1321 PLEETEELKRIIVDDTSTQWSKNMKHLTPSTLTQIDYNEKEKGAITQSPSLDCLTRSHI 1380

QY 1400 POANESPLPIAKVSSPPSIRPIYILTRVLFDQNSSHLPAASVYKKGDSGVQSSHFLQAGK 1459
DB 1381 POANESPLPIAKVSSPPSIRPIYILTRVLFDQNSSHLPAASVYKKGDSGVQSSHFLQAGK 1440

QY 1460 NNLSALITLMTGQREVGSLGTSATNSVYKVENTVLPKEDLPKTSKGVKELLPKVHI 1519
DB 1441 NNLSALITLMTGQREVGSLGTSATNSVYKVENTVLPKEDLPKTSKGVKELLPKVHI 1500

QY 1520 YQKOLFPPTETNSGSPGHLDLVEGSLLOGTGEGAIKWNENRPGKVPPLRVATBSSAKTPSK 1579
DB 1501 YQKOLFPPTETNSGSPGHLDLVEGSLLOGTGEGAIKWNENRPGKVPPLRVATBSSAKTPSK 1560

QY 1580 LLDPLAWNHNHVGTOIPKEEWSOEKSPKTAPEKKXDTILSLNACSNHAIAINNGQNP 1639
DB 1561 LLDPLAWNHNHVGTOIPKEEWSOEKSPKTAPEKKXDTILSLNACSNHAIAINNGQNP 1620

QY 1640 BIEVTWAKQGRTERLCSQNPVPLKHKHQREITRTTLQSDQEEIDYDDTISVEMKEDFDIY 1699
DB 1621 BIEVTWAKQGRTERLCSQNPVPLKHKHQREITRTTLQSDQEEIDYDDTISVEMKEDFDIY 1680

QY 1700 DDEENQSPRSPOKTRHYFIAAVERLWDYGMSSSHVLRNRAQSGSVQFKKVVQEBTD 1759
DB 1681 DDEENQSPRSPOKTRHYFIAAVERLWDYGMSSSHVLRNRAQSGSVQFKKVVQEBTD 1740

QY 1760 GSFTQPLVYRGELNEHLGLLGPVIRAEVEDNIMVTRPNOASRPYSYSSILSVIEEDQROGA 1819
DB 1741 GSFTQPLVYRGELNEHLGLLGPVIRAEVEDNIMVTRPNOASRPYSYSSILSVIEEDQROGA 1800

QY 1820 EPRKNFVKPNETKTYFWKVQSHVMAPTKDFDCKAWAYFSDVDLEKXDHSGLIGPLLVCHT 1879
DB 1801 EPRKNFVKPNETKTYFWKVQSHVMAPTKDFDCKAWAYFSDVDLEKXDHSGLIGPLLVCHT 1860

QY 1880 NTLNPAHQGRQTVQBFALFTTIFDEBTKSWYFTFENMERNCRAPCNQMEDPTPKENRPHA 1939

```

1661 NTLNPAFGQVTVQEFALFEFTTIDETKSWYFTENMERNCRAPCNIQWEDPTFKENYRFEA 1920
1940 INGYIMDTLPGLVMAODQRIWRLLSGNSNENIHSFSGHVFTVRKKBEYKMAVNLNLYP 1999
1921 INGYIMDTLPGLVMAODQRIWRLLSGNSNENIHSFSGHVFTVRKKBEYKMAVNLNLYP 1980
2000 GVFETVEMLPKAGIWRVCLAGEHLHAGMSTLFLVYSNKCOTPLQWASGHIRDFOITAS 2059
1981 GVFETVEMLPKAGIWRVCLAGEHLHAGMSTLFLVYSNKCOTPLQWASGHIRDFOITAS 2040
2060 GOYQOWAPKLARLHYSGSINAMSTKPFWSIKVDLLAPMIIHGIKTQGARQKPPSSVYISQ 2119
2041 GOYQOWAPKLARLHYSGSINAMSTKPFWSIKVDLLAPMIIHGIKTQGARQKPPSSVYISQ 2100
2120 FIIMYSLDGKKQWYRGNSGTILMVFGNVDSGIGKHNFNPPIIARYIRLHPHYSIRS 2179
2101 FIIMYSLDGKKQWYRGNSGTILMVFGNVDSGIGKHNFNPPIIARYIRLHPHYSIRS 2160
2180 TLRMELMGCDLNSCSYPLMGESKAISDAQITASSYFTNMFMATWSPSKARLHLQGRSNAR 2239
2161 TLRMELMGCDLNSCSYPLMGESKAISDAQITASSYFTNMFMATWSPSKARLHLQGRSNAR 2220
2240 PQVNNKEWLDQVDFQKTMKVTGVTTCGVKSLLTSMYVKZFLISSQDGHQWTLFFQNGKV 2299
2221 PQVNNKEWLDQVDFQKTMKVTGVTTCGVKSLLTSMYVKZFLISSQDGHQWTLFFQNGKV 2280
2300 KVFQGNQDSFTPVNSLDPPLTRYIRIHPQSWHVOIALRMEVLGCEAODLY 2351
2281 KVFQGNQDSFTPVNSLDPPLTRYIRIHPQSWHVOIALRMEVLGCEAODLY 2332

RESULT 12
US-10-187-319-6
GENERAL INFORMATION:
APPLICANT: Lollar, John S.
TITLE OF INVENTION: Hybrid Human/Animal Factor VIII
NUMBER OF SEQUENCES: 40
CORRESPONDENCE ADDRESS:
ADDRESS: Greenlee, Winner and Sullivan, P.C.
STREET: 5370 Manhattan Circle Suite 201
CITY: Boulder
STATE: Colorado
COUNTRY: USA
ZIP: 80303
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/187,319
FILING DATE: 27-Aug-2002
CLASSIFICATION: <Unknown>
PRIORITY INFORMATION:
APPLICATION NUMBER: US 09/523,656
FILING DATE: 2000-03-10
APPLICATION NUMBER: US 09/037,601
FILING DATE: 1998-03-10
APPLICATION NUMBER: WO PCT/US97/11155
FILING DATE: 1997-06-26
APPLICATION NUMBER: US 08/670,707
FILING DATE: 1996-06-26
ATTORNEY/AGENT INFORMATION:
NAME: Greenlee, Lorraine L.
REGISTRATION NUMBER: 27,894
REFERENCE/DOCKET NUMBER: 75-95X
TELECOMMUNICATION INFORMATION:
TELEPHONE: 303/499-8080
TELEFAX: 303/499-8089
TITLE: Sequence of the Murine Factor VIII cDNA
JOURNAL: Genomics
VOLUME: 16

```

```

PAGES: 374-379
DATE: 1993
RELEVANT RESIDUES IN SEQ ID NO: 6: FROM 1 TO 2319
SEQUENCE DESCRIPTION: SEQ ID NO: 6:
US-10-187-319-6

Query Match
Best Local Similarity 72.5%; Score 8827; DB 14; Length 2319;
Matches 1712; Conservative 232; Mismatches 364; Indels 54; Gaps 22;

QY 1 MQLRSTCFELCLARPCFSATRRYYIGAVELSDWYMQSD-LGELPVDARPPRPVPSPP 59
DB 1 MQLALFACPFELSLFPCSSAIRRYVIGAVELSNWYIQSDLLSVLHDSRFLPMSTSPF 60
QY 60 NTSVYVKKTLFVEFTDHLFNIAKPPPMGLLPTIOABVYTVVITLKNMASHVSLHA 119
DB 61 NTSIMYKTKVFEYKQDLFNIAKPP2PMGLLPTIWTVEVHTVITLKNMASHVSLHA 120
QY 120 VGVSYWKASGAEYDQCTSQREKEDKVPFGGSHYVYVQVLKENGPMASDPLCLYSYLS 179
DB 121 VGVSYWKASGAEYDQCTSQREKEDKVPFGGSHYVYVQVLKENGPMASDPLCLYSYLS 180
QY 180 HYDLVKDNLNSGLIGALLVCREGLAKEKTLKFIILFAVDEGKSWHSETPKSLMDR 239
DB 181 HYDLVKDNLNSGLIGALLVCREGLAKEKTLKFIILFAVDEGKSWHSETPKSLMDR 240
QY 240 DAASARAPKMHVTNGYVNRSLPGLIGHRKSVYVHVIHGMCTTPVHSIFLEGHTFLVRN 299
DB 241 DAASARAPKMHVTNGYVNRSLPGLIGHRKSVYVHVIHGMCTTPVHSIFLEGHTFLVRN 300
QY 300 HQASLEISPTITFLTAQTLMDLQGLLSCHISSHQHDMAYVKVDCSPBPOLRMK-N 358
DB 301 HQASLEISPTITFLTAQTLMDLQGLLSCHISSHQHDMAYVKVDCSPBPOLRMK-N 360
QY 359 NBEADYDDDDTDSMDVVRDDDDNSPSFIQIRSVAKKPKTWVHYIAAEEDWDYAPLV 418
DB 361 NEEMEDYDDDL-SEMDMFTLDYDSS-FIQIRSVAKKPKTWIHYIAAEEDWDYAPSV 418
QY 419 LAPDERSYKSVLNNGPORIGRKVKVPMAYTDTFTKTRAIQHESILGPLLYGEVD 478
DB 419 PTDNNGSYKSVLNNGPORIGRKVKVPMAYTDTFTKTRAIQHESILGPLLYGEVD 478
QY 479 TLLIIFKNQASRPYNIYPHGITDVRPLYSRRLPKGVKHLKOPPLPGEIFKYKWTVYED 538
DB 479 TLLIIFKNQASRPYNIYPHGITDVRPLYSRRLPKGVKHLKOPPLPGEIFKYKWTVYED 538
QY 539 GPTKSDPRCLTRYSSFFNMERDLASGLIGPLLCYKESVDQGNQIMSDKRNVLPSIF 598
DB 539 GPTKSDPRCLTRYSSFFNMERDLASGLIGPLLCYKESVDQGNQIMSDKRNVLPSIF 598
QY 599 DENRSWYLTENIQRFPLNPAGVQLEDPEFOASNTMHSINGVYVDSLOISVCLHEVAYWYI 658
DB 599 DENRSWYLTENIQRFPLNPAGVQLEDPEFOASNTMHSINGVYVDSLOISVCLHEVAYWYI 658
QY 659 LSVGAQTDFLSIFPSGYTFKKWVVEDTLTLFPGSETVFMMSMENPGLWLGCHNSDRK 718
DB 659 LSVGAQTDFLSIFPSGYTFKKWVVEDTLTLFPGSETVFMMSMENPGLWLGCHNSDRK 718
QY 719 RGMTHALLKVVSCDXTAGDYEDSYEDISAYLLSKNAIEPRSPQNSRHPSTROKQFNAI 778
DB 719 RGMTHALLKVVSCDXTAGDYEDSYEDISAYLLSKNAIEPRSPQNSRHPSTROKQFNAI 778
QY 779 TIPENDIEKTPDFAHRTMPKIQNVSSDILLMLLROS-PTEPHGLSLDLOEAKYEFSD 837
DB 779 TIPENDIEKTPDFAHRTMPKIQNVSSDILLMLLROS-PTEPHGLSLDLOEAKYEFSD 837
QY 838 DPSPCAIDSNNSLSEMTHTFRPQLHSGDMVFTSPSSQLRLNEXLGTATAELKKLPKV 897
DB 839 DHPNAIDSNNSLSEMTHTFRPQLHSGDMVFTSPSSQLRLNEXLGTATAELKKLPKV 897
QY 898 SSTNNLI-STIPSDNLAAGTNDTSSLPSPMPVHYDQDITTLFGKKSPLTBSGAPLS 956
DB 899 SSTNNLI-STIPSDNLAAGTNDTSSLPSPMPVHYDQDITTLFGKKSPLTBSGAPLS 956

```

QY 957 LSEENNDSKILPSGLMNSOESSGKGVSVSTESGRLFKGRAGHPALLTKONALFKVSI 1016
DB 959 ASZENSNDILSTLMYSESI PRDNLISDIENDRLREFFHGIALLTKONTLFDONVSL 1018
QY 1017 LKTNKTSNNAIRKTHIDGPSLLENSPSVQWQ--ILESDETFKVKVPLIHDRMLMDKNA 1075
DB 1019 MKTKNTYHSTTNEKLHESPT--SIENSTDLQDAILKYNSEIQEVTALIHDTGLAGKNS 1077
QY 1076 TALRLNHXNKTTSSKNMEXVQKGGPIPPDQNDPMDFKMLFIPESARWIORTHGN 1135
DB 1078 TYLRNLHMLNRITSTKNOIFHRKDEPIPODEINTIPFSKMLFLESSESNFKTINGNN 1137
QY 1136 SLNSGQSPKQIVLSLGPESKVEGQNFLEKKNVGVKGEFTKDVGLKEMVPPSSNLF 1195
DB 1138 SLNSEGEHSPKQVLYLMPFKYKYNQSFLEKKNVYEQDGFKNIGKDMAFPHNNSIFL 1197
QY 1196 TMLDNHNNHNTNQEKKIOEIEKKETLLOENVLPQIHVITGTRKPMKNLFLSTRQNV 1255
DB 1198 TILSNVHENGRENQEKNIQEEIE--KEALTEEKVVLQVHEATGSKNFKLIDILITGRONI 1256
QY 1256 EGSYDGAAYAVLODFPSLNDSTNRTKKHTAHSK--KGBEENLEGLGNQTKQIVEKYACT 1313
DB 1257 --SLYEVHVPVLQNTISNNSTNTVQIHMHEHFPKREKDKETNSEGLVNTREKVNKY--- 1311
QY 1314 TRISPTSQNFVTOFSKRALQKQFRLPLEBETELEKRIIVDDTSTOKSKMKHLPSTLTQ 1373
DB 1312 -----PSQKNITTOFSKALGQFRL-----STQWLKTINGSTQIILQ 1349
QY 1374 IYNEKEKGAITOSPISDCLTRSHSIPQANRSPPIAKVSSPFSIRPYLTVLRFVLFODNS 1433
DB 1350 IDHSEKMKFIFKSSLSDS--SVIKSTQTNSSDSHIVKTSAPP--PIDLKRSPFONKES 1405
QY 1434 HLPAAZY---RKDSSGOESHFLOGAKKNLSLAILLEMTGQREVGSIGTSATNSV 1489
DB 1406 HVQASSYIYDFKTSRIQESNNFLKETKINPSTAILPWNPIFIDGKTSPCKSTNSV 1465
QY 1490 TYKKVENTVLPKDPLEKTSKVELLPKVHYQKDLFPETTSNGSPGHLDVIGSLQGTB 1549
DB 1466 TYKKRENIIFLKTPEESKTELLPQVSOBEIILPTSHGSPGHNLMEVFLOKIQ 1525
QY 1550 GAIKNEANRPGVPLFRVATISSAKTPSKLDPLAWDNYHGTQIPKEBWKQSEKSPKT 1609
DB 1526 GPTKWNKAKRHGB--SICKTSSSKNTRSKLLNHHAWDVHYAAQIPKDMWKSKEKSPBII 1583
QY 1610 AFKKKOTILSLNACENSHAIANEQNKZEEIETWAKQGRTERELCSQNPVILKHHORI 1669
DB 1584 S-KQEDTILSLPHGNHSHIGA--NEKQNPQRETTWVKQGTQRTCSQIPPVILKHHOREL 1642
QY 1670 TRTTIQSDQEEIDYDDTISVENMKEDFDIYDEQNSPFSQKTRHYFIAAVERLWDYG 1729
DB 1643 --SAFQSECEATDYDAITIE--TIEDFDIYSEDIKQGPSFOCKTBYHYFIAAVERLWDYG 1699
QY 1730 MSSSPHVLNRAQSGSVPOFKVVPQFTDGGPTQPLRYGELNEHGLLGPYIRAEVDN 1789
DB 1700 NESTS--HVLNRYQSDNVPOFKVVPQFTDGGPSQPLRYGELNEHGLLGPYIRAEVDN 1758
QY 1790 IAVTFNQASRPYSFYSSILSYEEDORQGAERKPNFVKNETKYPFKVQHMNAPTKQSF 1849
DB 1759 IAVTFNQASRPYSFYSSILSYEEDOR--GEBPRNFVKNETKYPFKVQHMNAPTKQSF 1817
QY 1850 DCKAWAYFSDVLEKDVHSLGIPLIVCHTNTLNPARGSQVTVQEPALFTTFIDETKSWY 1909
DB 1818 DCKAWAYFSDVLEKDVHSLGIPLIIICANTLNPARGSQVTVQEPALFTTFIDETKSWY 1877
QY 1910 FTENKERNACPNIOEMDPTFKENTRFAHNGYIMDTLPGLVMAQDQIRWYLLSMGN 1969
DB 1878 FTENKRNCKTCFNFQEMDPTLKENYRFAHNGYIMDTLPGLVMAQDQIRWYLLSMGN 1937
QY 1970 ENIHSIHFSCHVFTFKKKEYKVALNLYGVPETVEMLPFSKAGIWRVSLIGELHACM 2029
DB 1938 ENIQSIHFSCHVFTFKKKEYKVALNLYGVPETVEMLPFSKAGIWRVSLIGELHACM 1997

QY 2030 STLFVYVSNKCOITPLGMASGHIRDFQITASGOYGQWAPKLARLHYSGSINAMSTKEPPSW 2089
DB 1998 STLFVYVSKCOITPLGMASGIRDFQITASGHYGQWAPNLAELHYSGSINAMSTKEPPSW 2057
QY 2090 IKVDLLAPMIHIGIKTOGARQKFSLSYISQFIIMTSLDGKKWQTYRGNSTGTLMVFFGNV 2149
DB 2058 IKVDLLAPMIHIGIKTOGARQKFSLSYISQFIIMTSLDGKKWLSYQGNSTGTLMVFFGNV 2117
QY 2150 DSSGIXHNIENPPIIARIYIRLHPTHYSTRSTLMELMGCDLNSCMPLGMEKSAISDAOI 2209
DB 2118 DSSGIXHNSFNEPIIARIYIRLHPTHYSTRSTLMELMGCDLNSCIPLGMEKSVISDOI 2177
QY 2210 TASSYFTNMFATWSPSKARLHLOGESNARPOVNNPKWLOYDFOXTMKVGTGVTGQVKS 2269
DB 2178 TASSYFTNMFATWSPSQARLHLOGHTNARPOVNDPKQWLQVDLQKTMKVTGIIITQGVKS 2237
QY 2270 LITSYVKEFLISSQDGHQWTLFTQNGKVKVQGNQOSTFTFVNSLDPPLTRYLRIHP 2329
DB 2238 JFTSMFVKEFLISSQDGHHTQIILYNGKVKVQGNQOSTFTMNSLDPPLTRYLRIHP 2297
QY 2330 QSWHQAIALRMEVLGCEAQQDY 2351
DB 2298 QIWEHQIALRLEILGCEAQQDY 2319

RESULT 13

US-10-131-510A-6
; Sequence 6, Application US/10131510A
; Publication No. US20030166536A1
; GENERAL INFORMATION:
; APPLICANT: Lollar, John S
; TITLE OF INVENTION: Modified Factor VIII
; FILE REFERENCE: 75-95J
; CURRENT APPLICATION NUMBER: US/10/131,510A
; CURRENT FILING DATE: 2002-07-10
; PRIOR APPLICATION NUMBER: U.S. 09/315,179
; PRIOR FILING DATE: 1999-05-20
; PRIOR APPLICATION NUMBER: U.S. 09/037,601
; PRIOR FILING DATE: 1998-03-10
; PRIOR APPLICATION NUMBER: U.S. 08/670,707
; PRIOR FILING DATE: 1996-06-26
; PRIOR APPLICATION NUMBER: PCT/US97/11155
; PRIOR FILING DATE: 1997-06-26
; PRIOR APPLICATION NUMBER: PCT/US94/13200
; PRIOR FILING DATE: 1994-11-15
; PRIOR APPLICATION NUMBER: U.S. 08/212,133
; PRIOR FILING DATE: 1994-03-11
; PRIOR APPLICATION NUMBER: U.S. 07/864,004
; PRIOR FILING DATE: 1992-04-07
; NUMBER OF SEQ ID NOS: 40
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 6
; LENGTH: 2319
; TYPE: PRT
; ORGANISM: Mus musculus
US-10-131-510A-6

Query Match 71.1%; Score 8827; DB 14; Length 2319;
Best Local Similarity 72.5%; Pred. No. 0;
Matches 1712; Conservative 232; Mismatches 364; Indels 54; Gaps 22;
QY 1 MOELSTCFCLLFCPSATRYVYLGAVELSWDYMQSD--LCELVDARFPPRPVKSPFF 59
DB 1 MQIALFAFFCLLNFPCSSAIRYVYLGAVELSWYIQSDLLSVLETDTSRFLPRMSTSPFF 60
QY 60 NTSVYVYKTLFVEFTDHLFNIAKPRPPWGLGPTIOAEVYDVTVTITKNMASHPVSLHA 119
DB 61 NTSIMYKKTVFVEYKQDLFNIAKPRPPWGLGPTIWEVHTVITILKNMASHPVSLHA 120
QY 120 VGVSVYKASEGAEYDQTSQREKEDKVPFGSGSHYVWQVLKENGPMASDPCLTYISLS 179
DB 121 VGVSVYKASEGEYEDQTSQMEKEDKVPFGSGSHYVWQVLKENGPMASDPCLTYISYS 180

QY 180 HYDLVKDLNSGLIGALLVCREGLSLAKEKTQTTLHKFILLFAVEDEGKSHSETKNSLMQDR 233
 Db 181 HYDLVKDLNSGLIGALLVCKEGLSKERTOMLYQFVLFAVEDEGKSHSEINDSYTQSM 240
 QY 240 DAASAAWPMHTVNGYVNRSLPGLIGCHRXKSYVMEVIGMTTPEVHSLFEGHTFLVRN 299
 Db 241 DSASAADWPMHTVNGYVNRSLPGLIGCHRXKSYVMEVIGMTTPEHSLFEGHTFFVRN 300
 QY 300 HQASLEISPIITPLACTLMDLGOFLLSCHLSSHOHGMWAAVVKVDSCEPSPQLRMK-N 358
 Db 301 HQASLEISPIITPLACTLMDLGOFLLSCHLSSHOHGMWAAVVKVDSCEPSPQWQKNN 360
 QY 359 NEAEADYDLDLSEMDVVRFDNDSPFIQIRSAVKHPKTWVHYIAAEEDWDYAPLV 418
 Db 361 NEEMEDYDDLLY-SEMDMFTLDYDSSP-FIQIRSAVKHPKTWVHYIAAEEDWDYAPLV 418
 QY 419 LAPDRSYKSOYLNCPORIGRKYKVRPMAYTDETFKTRBAIQHSGILGPLLGEVGD 478
 Db 419 PTDNGSYKSOYLSNGPHRIGRKYKVRFIAYTDETFKTRBAIQHSGILGPLLGEVGD 478
 QY 479 TLLIIFKNOASRPYNIYPHGITDVRPLYSRRLPKGVKHLKDPILPGBETPKYKWTVTVED 538
 Db 479 TLLIIFKNOASRPYNIYPHGITDVRPLYSRRLPKGVKHLKDPILPGBETPKYKWTVTVED 538
 QY 539 GPTKSDPRCLTRYSSFVNMEXDLASGLIGPLLI CYKESVDQRGNQIMSDXRNVLFSYF 598
 Db 539 GPTKSDPRCLTRYSSFINFERDLASGLIGPLLI CYKESVDQRGNQIMSDXRNVLFSYF 598
 QY 599 DENRSVLENTORPLNPAGVLEBPBPQASIMHSINGYVDSLOLVCHEVAWYVI 558
 Db 599 DENQSWYITENORPLNPAGVLEBPBPQASIMHSINGYVDSLOLVCHEVAWYVI 558
 QY 659 LSIGAOTDLSVFFSGYTPKHVMYEDTLTLFPFSGETVFMSENPGMLTILGCHNSDFRN 718
 Db 659 LSVGAOTDLSVFFSGYTPKHVMYEDTLTLFPFSGETVFMSENPGMLTILGCHNSDFRN 718
 QY 719 RGMNTALLKVSXCDKNTGDYVEDYEDISAYLISKNNALBPRSPQNSRHPSTROKFNAT 778
 Db 719 RGMNTALLKVSXCDKNTGDYVEDYEDIPQLVNNENNVIDPRSPFQNTNHPNTRKXKFKDS 778
 QY 779 TIPENDIEKTDPFPAHRTMPKQONVSSSLLMLLRQS-PTPHGLSLDLOERAKYTFSD 837
 Db 779 TIPKNDMEKLEPOFEIEAEMLKVCQSVSDMLLGGOSHPTPEGLFLSDGQEAIEAIDH 838
 QY 838 DESPGAIDSNNSUSEMTHFAPQLHSGDMVFTPESGLOQLRNEKLGTTAATELKLDLDFV 897
 Db 839 DHPNALDSNEGSPSKVTQAPESHSEKTVFTPPQGLQLRSNKSLETTIEVKKKLGLQV 898
 QY 898 SSTSNMLI-STIPSDMLAGTUNTSILGPPSPMPPHYVDSQDITLFGKKSPLTESGGPLS 956
 Db 899 SLPFSNLMTTTLISDLNKATFEKTDSSGFPDMPPVHSSKSLSTTAFGKKAYSLVGSHVPLN 958
 QY 957 LBRENNDSKILBSGLNQSBSKGNVSVSTESGRFLPKGRAHPALLITKONALFKVYSIL 1016
 Db 959 ASEENSNDLSTLMYSQESLPRDNLSTENDRELLREKXPHGIALITKONLIFKONVSL 1018
 QY 1017 LXTNKTNNATNRKTHIDGPSLLIENSPSVWQN-ILESDTEPKKVTPLIHDRMLDKNA 1075
 Db 1019 MKNTKYNESTNEKLTHTSPT-SIENSTTDLQDAILKVNSEIQEVETALIHDDOTLLGKNS 1077
 QY 1076 TALRIAHMSKNTTSSKNNEMVQOKKGGPIPPDAQNDMDSPFKMLFPESARWQRTHGKN 1135
 Db 1078 TYLRHMLNMTTSTKNTKOTIFHRKDDPIPODEENTIMPFSKXLFLESNNWFKKTGN 1137
 QY 1136 SLNSGQGPSKQOLVSLGPEKSVGQPLSEKNNKVVWVGKGFETDVLGKEMVFPSSNLF 1195
 Db 1138 SLNSEQHSKQOLVYLKFKYKVNQSPFLSEKNNKVTVEQDFTNIGLKDWAFPHNMSIFL 1197
 QY 1196 TNLNLEHNTHNQEKKIQEIEKKTFLIQENYVLPQIHTVTCFKNFMKNLFLSTRQNY 1255
 Db 1198 TTLNHNHGRHNRQEKNIQIEIE-KZALIEBKVLPQVHEATSKNFKLDIILGTRQNI 1256
 QY 1256 EGSYDGAIVLQDPRSLNDSNTRTKHTAHFSK--KGSEENLEGLGNQTKQIVEKYACT 1313

Db 1257 --SLYEVHPVLQNTS--NNSINTVQIHEHFFKRRDKETNSEGLVNKTRVMKY --- 1311
 QY 1314 TRISNTSQOHPVTCRKRALKOPRLPLEETELEKRIIVDDTSTOWSKMKHLFSTLTQ 1373
 Db 1312 -----PQKNIITQSRKALQOPL-----STQWLKTINGSTQCIKQ 1349
 QY 1374 IDYNEKEGALTQSPLSDCLSTRSHSIPOANRSPLIPIAKVSPSPSIRPYLTRVLFDQNS 1433
 Db 1350 IDHSEMKKFTTKSLSDS-SVIKSTTQTNSSDSHIVKTSAPP---PIDLKRSPONKFS 1405
 QY 1434 HLPAAZY----RKXDSGVQESHFIQGAKNKLSLAILTLEMTQDQREVGSLGTSATNSV 1489
 Db 1406 HVQASSYIVDFKTESR-QESNNFLKTKINNPSSAILPWNFIIDQKFTSPGKNTNSV 1465
 QY 1490 TVYKVENYVLPKPLPKTSGKVELLPKVIHOKDLPPTETNSGSPHLLDVBGSLQGT 1549
 Db 1466 TYKGENIIFLPTLPBESGKIEJLPQVIOBEELTPTETSHGSPHLLDVBGSLQGT 1525
 QY 1550 GAIKWEANRPGKVPFLRVATESSAKIPSKLIDPLAWNHYGTQIPKBEWKSQESPEKT 1609
 Db 1526 GPTKNAKKEGE--SIKKTIBSSANTSKLNNHENDYHYAAQIPKOWMKSEKSPBII 1583
 QY 1610 APKKDITLUSNACHSNHAIANINQONKPE-EVTWAKOGRTERLCSQNPVVLKXHQREI 1669
 Db 1584 SIKQEDTILSLAPHGNSHIGA-NEKQNPQRETTWVKQGTQRTCSQIPTPVLKHQREI 1642
 QY 1670 TTTTLOSQEBIDYDDTISVEMKKEDFDIYDDENSPRSFOKTRHYPIAAVEBLWDY 1729
 Db 1643 --SAQSEQATYDDAITIE-TIHDFDIYSSDIKQGRPSFOKTRHYPIAAVEBLWDY 1699
 QY 1730 MSSSHVLNRAQSGSVPOFKKVPQETDGSFTQPLRGELNEHLGLIPYIRABVEDN 1789
 Db 1700 XSTS-HVLNRYQSDNVPOFKKVPQETDGSFTQPLRGELNEHLGLIPYIRABVEDN 1758
 QY 1790 IWTYRNQASRPYSYSSLSIYBEEORQGAERKPNFVKPNETKTYFKVQOHHMATKUBF 1849
 Db 1759 IWTYRNQASRPYSYSSLSIYBEEORQGAERKPNFVKPNETKTYFKVQOHHMATKUBF 1817
 QY 1850 DCKWAYFSDVLEKDVHSGILGPLLCHNTLNPAGHGFQVTVQEFALFTTIDFTKSY 1909
 Db 1818 DCKWAYFSDVLEKDVHSGILGPLLCHNTLNPAGHGFQVTVQEFALFTTIDFTKSY 1877
 QY 1910 FTENMERNCRAPCNQIMEDPTFKENYRPHAINGYIMDTLPGLVMAQDQIRWYLLSMGN 1969
 Db 1878 FTENMERNCRAPCNQIMEDPTFKENYRPHAINGYIMDTLPGLVMAQDQIRWYLLSMGN 1937
 QY 1970 ENHSHIFSGHYFTYRKKEEYXVALYNLYPGYFVEMLPKAGIWRVECLIGEHLQAGN 2029
 Db 1938 ENIQSHIFSGHYFTYRKKEEYXVALYNLYPGYFVEMLPKAGIWRVECLIGEHLQAGN 1997
 QY 2030 STLFLVYSNKCOTPLGMASGHIRDFQITASGOYGQWAPKLARLHYSGSINAWSTKEPPSW 2089
 Db 1998 STLFLVYSNKCOTPLGMASGHIRDFQITASGOYGQWAPKLARLHYSGSINAWSTKEPPSW 2057
 QY 2090 IKVDLLAPMIHSGIKTQGARQFSSLYISQFIIMTSLDGKKWQTYRGSTGLMYPFGNV 2149
 Db 2058 IKVDLLAPMIHSGIKTQGARQFSSLYISQFIIMTSLDGKKWQTYRGSTGLMYPFGNV 2117
 QY 2150 DSGTGKHNIFNPPIIARVIRLPHTHYSIRSTLRMELMGDLNCSMPLGMBESKASDAQI 2209
 Db 2118 DSGTGKHNIFNPPIIARVIRLPHTHYSIRSTLRMELMGDLNCSMPLGMBESKASDAQI 2177
 QY 2210 TASSYFTNMFAWSPSKARLHLQGRSNARWPOVNNPKWLOVDQRTKWTGVTITQGVKS 2269
 Db 2178 TASSYFTNMFAWSPSKARLHLQGRSNARWPOVNNPKWLOVDQRTKWTGVTITQGVKS 2237
 QY 2270 LFTSMYVKEFLISSQDGHQWTLFTFQNGKVKVQGNQSFPTVNSLDPPLLTRYLIRHP 2329
 Db 2238 LFTSMYVKEFLISSQDGHQWTLFTFQNGKVKVQGNQSFPTVNSLDPPLLTRYLIRHP 2297
 QY 2330 QSWHCHIALRMEVLGCEAQDLY 2351

Db 2298 QIWEHQIALREILGCRQQY 2319

RESULT 14

US-10-187-319-37

Sequence 37, Application US/10187319

Publication No. US20030068785A1

GENERAL INFORMATION:

APPLICANT: Lollar, John S.

TITLE OF INVENTION: Hybrid Human/Animal Factor VIII

NUMBER OF SEQUENCES: 40

CORRESPONDENCE ADDRESS:

ADDRESSEE: Greenlee, Winner and Sullivan, P.C.

STREET: 5370 Manhattan Circle Suite 201

CITY: Boulder

STATE: Colorado

COUNTRY: USA

ZIP: 80303

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent In Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/10/187,319

FILING DATE: 27-Aug-2002

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 09/523,656

FILING DATE: 2000-03-10

APPLICATION NUMBER: US 09/037,601

FILING DATE: 1998-03-10

APPLICATION NUMBER: WO PCT/US97/11155

FILING DATE: 1997-06-26

APPLICATION NUMBER: US 08/670,707

FILING DATE: 1996-06-26

ATTORNEY/AGENT INFORMATION:

NAME: Greenlee, Lorraine L.

REGISTRATION NUMBER: 27,894

REFERENCE/DOCKET NUMBER: 75-95K

TELECOMMUNICATION INFORMATION:

TELEPHONE: 303/499-8080

TELEFAX: 303/499-8089

INFORMATION FOR SEQ ID NO: 37:

SEQUENCE CHARACTERISTICS:

LENGTH: 2133 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

SEQUENCE DESCRIPTION: SEQ ID NO: 37:

US-10-187-319-37

Query Match 65.0%; Score 8068; DB 14; Length 2133;

Best Local Similarity 67.0%; Fred. No. 0;

Matches 1577; Conservative 208; Mismatches 344; Indels 226; Gaps 17;

QY 1 MOELSTCFLLCFRCFATRRYVLGAVELSDYMQSD-LGELPVDARFPRVPKSPFF 59

Db 1 MOELSTCFLLCFRCFATRRYVLGAVELSDYMQSD-LGELPVDARFPRVPKSPFF 60

QY 60 NTSVYKKTFLVFETDQLFNSIAKPRPPWGLGPTIQAEVDTVITLKNMASHFVSLHA 119

Db 61 GPSVLYKKTFLVFETDQLFNSIAKPRPPWGLGPTIQAEVDTVITLKNMASHFVSLHA 120

QY 120 VGVSVMKASEAGVDDQTSOREKEDKVPFGSGSHYVQVLKENGCPMADDPCLITYSL 179

Db 121 VGVSVMKASEAGVDDQTSOREKEDKVPFGSGSHYVQVLKENGCPMADDPCLITYSL 180

QY 180 HVDVLKDLNSGLIGALLVCREGLAKEKTQTLHKPFIILFAVDFDEGKSWHSETQNSLMODR 239

Db 181 HVDVLKDLNSGLIGALLVCREGLAKEKTQTLHKPFIILFAVDFDEGKSWHSETQNSLMODR 240

QY 240 DAASARAWFKKHTVNGYVNRSLPGLIGCHRSVYWHVIGMGTTPVHSGIFLEGHTFLVRN 299

241 DPAPARAQPAHHTVNGYVNRSLPGLIGCHRSVYWHVIGMGTTPVHSGIFLEGHTFLVRH 300

300 HROASLEISPIITLTAOTLLMDIGQELLCHISHSHOHGMEAYVYKVDSCPEEPOLRMKN 359

301 HROASLEISPIITLTAOTLLMDIGQELLCHISHSHOHGMEAYVYKVDSCPEEPOLRMKN 360

360 EBAEDYDDDLTSEMDFVRFDNDSPSFIQIRSVAKKHPTKTVHYIAAEEEDMDYAPLVL 419

361 EE-EDYDDNLYDSMDVVRDLGDVDSFFIQRISVAKKHPTKTVHYIAAEEEDMDYAPLVL 419

420 APDRSYKSYLNGPQIRIGRYKVKRFMAYTDETFKTRAEAIQHESGILGPLYGEVGT 479

420 SPDRSYKSYLNGPQIRIGRYKVKRFMAYTDETFKTRAEAIQHESGILGPLYGEVGT 479

480 LLIIIFKNQASRPYNIYPHGITDVYRSLRPLPGVGHLDKDPILPGEIFKVKTVTVEDG 539

480 LLIIIFKNQASRPYNIYPHGITDVYRSLRPLPGVGHLDKDPILPGEIFKVKTVTVEDG 539

540 PTKSDPRCLTRYSSFVNMRDLASGLIGPLICYESVDQGNQIMSKRNVLFSVFD 599

540 PTKSDPRCLTRYSSFVNMRDLASGLIGPLICYESVDQGNQIMSKRNVLFSVFD 599

600 ENRSWYLTENIQRLPNPAGVQLEDPPFOASNMHESINGYVFDLSQLSVCLHEVAYWYL 659

600 ENRSWYLTENIQRLPNPAGVQLEDPPFOASNMHESINGYVFDLSQLSVCLHEVAYWYL 659

660 SIGAQTDFLSVFGSYTFKHKMYVEDTLTPPSGETVFMENPGLWILGCHNSDNRN 719

660 SVGAQTDFLSVFGSYTFKHKMYVEDTLTPPSGETVFMENPGLWILGCHNSDNRN 719

720 GMTALLKVVSCDKNTGYEDSYEDISAYLLSKNNAIEPRSFQNSRHPSTROKQFVATT 779

720 GMTALLKVVSCDKNTGYEDSYEDISAYLLSKNNAIEPRSFQNSRHPSTROKQFVATT 779

780 IPENDIEKTPWFAHRTPMKIQNVSSDDLMLLRQSPTHGLSLDLSQAKYTFSDDP 839

780 SPEDDVE-LDPQSGERTQALELSVPSGDSMLLGNPAPHGSSSSDLQEARNE--ADDY 836

840 SPGAIDSNNSLSBMTHERPOLHSGDMVFTPESSQLRLNEKLGTAAATELKKLDFKVS 899

837 LPGARENTAPSAARLRPELHSAEVLTPP-----EK-----EUKLDSKMS 882

900 TSNMLIS--TIPSDNLAAGTNTSSLGPPSPMPHYDVSQDLOTTLFGKSSPLTEGGPLSL 957

883 SSDLLKTSPIIPSDTLGAEHTHSLGPPHPPQVNFNRSQLGAILGKNSHPIGAGVPLGS 942

958 SEENNSKLLSEGLMNSQBSGWKNSVSTSGRLFKGKRAHGPALLTKDNALFKVSI 1017

943 TEEDH-----ESSLGENVSPVSDGIFEKERAHGPASLTAKDDVLFKVNISLV 989

1018 KTNKTSNNSATNRKTHIDGPSLLTENSVPVQNLIESJTEPKKVTPLIHDRMLDKNATA 1077

990 KTKARVILKTNRIKIHDDAALUTENASA-----TFMDKNNTTA 1028

1078 LRLNEMNKTTSSKMEMVQCKEGPIPPDAQNPDMSFFKMLFLPESARWIQTHGKNSL 1137

1029 SGLAHVN-----MIKGLGKNPL 1047

1138 NSGOGSPKQVLSLGPKEVGEQNFLEKKNVUVKGEFTKDVGLKEMVFPSSRNFLTN 1197

1048 SSERGPSPELLTSSSGSKVKGSGQGRIRAVAEELSKG---KEMMLFNSLTLTN 1104

1198 LDNLHNNHNOEKIOHEIEKKTETLIQENVLPIQHTVTGTKNFMKNLFLLSRQVVEG 1257

1105 SADVQGNTHSQGKSRSEMERREKLVQEKVDDLPQVTAITGKTNFLRNIHQSTEPSVEG 1164

1258 SYCGAVAPVLQDFRSLNDS--NRTKKHTAHSKKEENLEGLNQCTQIVEKYACTTIRIS 1317

1165 FDGSHAPVPCDSLSLNSAERAEATHIAHSAIREAPLEAPGNRT----- 1210

1318 PNTSQNFVTORSKALKQFRLPLEETELEKRIIVDDTISTQWSKMKHLTPSTLTQIDYN 1377

Db 1211 -GPGPRSAVPRVKQSLQIRLPLEEIKPGRGVNLNATSRWS----- 1252
Qy 1378 EKEGAIQTSPSLDCLTRSHSIPQANRSLPIAKVSSPPSIRPIYLTRVLFDQNSHLPA 1437
Db 1253 ----- 1252
Qy 1438 ASYRKXDSGVQSSSHFLOCAKNNLSLAILTLEMTGDQREVSLGTSAPNSVYTKVENT 1497
Db 1253 -----ESSPILOCAKNNLSLPTLLEMGAGGCKISALGKSAAGPLASGKLEKA 1301
Qy 1498 VLPKPDLPKTSKVELLPVAVHIYQKDLPTETSNOSPGHLDIVESLLOCTEGAKWNEA 1557
Db 1302 VLSSAGLSASAKAEFLPVVRHEDLLFQKTSNVSCAHDGLQIFLQKTRGPNLANKV 1361
Qy 1558 NRPFGVPFLRVATSSAKTSPSKLLDPLAWDNHYGTQIPKEEMKSOEKSPKTAFAKKOTI 1617
Db 1362 NRPFG-----RTPSKLLGP-----PMPK-EWESLEKSPKSTALATKDII 1398
Qy 1618 -LSLNACENHAIAINEQNKPEIEVTWAKQRTERICSONPPVLKRHOREITRTLOS 1676
Db 1399 SLPLORHESNHSIAKNEQABETOREAAWTKQGGPRLCAPKPPVLRHQRDISLPTFOP 1458
Qy 1677 DOEE-DYDDTISVEMKEDFDIYDENOSPRSFOKTHYFIAAVERLWDYGMSSPHV 1736
Db 1459 BEKMDYDDIFSTETKGEDFDIYGEDENQDPRSFOKTRHYFIAAVEQWLDYGMGESPPA 1518
Qy 1737 LNRQAQSGSVPOFKKVVFQEFDTGSGFTOPLRYGELNEHLGLGPIYTRAEVDNIMVTFN 1796
Db 1519 LNRQAQNGVPRPKVVPREFADGSGTQPSYRGELNKLGLGPIYTRAEVDNIMVTFN 1578
Qy 1797 QASRPYSFSSLSIIEEDQOAGAPRKNFVKNETKYTWKVOHMAPTKDSFDCWAY 1856
Db 1579 QASRPYSFYSSLSIYDDEQOAGAPRKNFVQNETRYTWKVOHMAPTKDSFDCWAY 1638
Qy 1857 FSDVLEKDVHSLGILPVLVCHTNTLNPAHGQVTVQEPALFTTIDETKSWFYFENMER 1916
Db 1639 FSDVLEKDVHSLGILPVLVCRANTLNAAGHQVTVQEPALFTTIDETKSWFYFENMER 1698
Qy 1917 NCRAPCNIQMEDPTFKENYRPHAINGYIMDTPLGLVMAQDQIRMYLLSGMGENIHSIH 1976
Db 1699 NCRAPCHLQVEDPTLKENYRPHAINGYIMDTPLGLVMAQDQIRMYLLSGMGENIHSIH 1758
Qy 1977 FSGHYFTVPEKKKEYKALYNLPGVPEFTVEMLPKAGIWRVCECLIGEHLHAGMSILFLVY 2036
Db 1759 FSGHYFVPEKKKEYKAVNLNPGVPEFTVEMLPKAGIWRVCECLIGEHLHAGMSILFLVY 1818
Qy 2037 SNKCOPTLGMASGHIRDFOITASGOYGOWAPKLARLHYSGSINANSTKPFPMKVDLLA 2096
Db 1819 SKECQAPLGMASGRIRDFOITASGOYGOWAPKLARLHYSGSINANSTKPFPMKVDLLA 1878
Qy 2097 PMTHGKTQGARQKPSLIYISQFIIMYSLDGGKWTYRGNSTGTLVMVFGNVDSGKIXH 2156
Db 1879 PMTHGIMTQGARQKPSLIYISQFIIMYSLDGRNMQSYRGNSTGTLMVFFGNVDASGKIXH 1938
Qy 2157 NFNPNPIARYIRLPHTHYSIRSLTRELMLGCDLNSCMLPGWESKAIASDAQITASSYFT 2216
Db 1939 NFNPNPIARYIRLPHTHYSIRSLTRELMLGCDLNSCMLPGWESKAIASDSQITASSHUS 1998
Qy 2217 NMFATWSPKARLHLQGRNAMEPOVNNPREMLQVDFQTKMKVTGVTTOGVKSLITSMVY 2276
Db 1999 NMFATWSPKARLHLQGRNAMEPRVSSAEEMQLVDLQTKVTKVTGVTTOGVKSLITSMVY 2058
Qy 2277 KEPLISSQDGHOWITLFFQNGKVKVFGNQDSTFPVNSLDPLPLTRYLRIRHPSQWVHOI 2336
Db 2059 KEPLVSSQDGRWITLFLQDGHITKVFQNGQDSTFPVNSLDPLPLTRYLRIRHPSQWVHOI 2118
Qy 2337 ALRMEVLGCEAODLY 2351
Db 2119 ALRLEVLGCEAODLY 2133

RESULT 15

US-10-131-510A-37

; Sequence 37, Application US/10131510A
; Publication No. US20030166536A1
; GENERAL INFORMATION:
; APPLICANT: Lollar, John S
; TITLE OF INVENTION: Modified Factor VIII
; FILE REFERENCE: 75-95U
; CURRENT APPLICATION NUMBER: US/10/131,510A
; CURRENT FILING DATE: 2002-07-10
; PRIOR APPLICATION NUMBER: U.S. 09/315,179
; PRIOR FILING DATE: 1999-05-20
; PRIOR APPLICATION NUMBER: U.S. 09/037,601
; PRIOR FILING DATE: 1998-03-10
; PRIOR APPLICATION NUMBER: U.S. 08/670,707
; PRIOR FILING DATE: 1996-06-26
; PRIOR APPLICATION NUMBER: PCT/US97/11155
; PRIOR FILING DATE: 1997-06-26
; PRIOR APPLICATION NUMBER: PCT/US94/13200
; PRIOR FILING DATE: 1994-11-15
; PRIOR APPLICATION NUMBER: U.S. 08/212,133
; PRIOR FILING DATE: 1994-03-11
; PRIOR APPLICATION NUMBER: U.S. 07/864,004
; PRIOR FILING DATE: 1992-04-07
; NUMBER OF SEQ ID NOS: 40
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 37
; LENGTH: 2133
; TYPE: PRT
; ORGANISM: Sus scrofa
US-10-131-510A-37

Query Match 65.0%; Score 8068; DB 14; Length 2133;
Best Local Similarity 67.0%; Pred. No. 0;
Matches 1577; Conservative 208; Mismatches 344; Indels 236; Gaps 17;
Qy 1 MOELSTCFECLLAFCSATRRYVLGAVLSWDQMOSD-LGELPVDARFEPVRVSPKF 59
Db 1 MOELSTCVFCLLPGLFSAIRRYVLGAVLSWDYRQSELRRELVHVDTRFPATAPCALPL 60
Qy 60 NTSVYVYKTLFVEFTDHLFNIAKPRPPMWMGLLGPITQAEVYDVVITLKNWASHPVSLHA 119
Db 61 GFSVLYKKTZVFEFTDQLFVSARPPPMWGLLGPITQAEVYDVVITLKNWASHPVSLHA 120
Qy 120 VCVSYWVKSAGEAYDDQTSOREKEDDKVPFGSHTYVWQVLKENGPMASDPLCLITYSL 179
Db 121 VCVSWFKSSEGAAYSDHTSOREKEDDKVLPKGSQTYVWQVLKENGPTASDPPCLITYSL 180
Qy 180 HYDLVKDLSNGLIGALLVCREGLAKETQTLHKETILLFAVDFDEGKSWHSETKNSLMODR 239
Db 181 HYDLVKDLSNGLIGALLVCREGLTRETQNLHEFVLIFPAVDFDEGKSWHSAENDSWTRAM 240
Qy 240 DAASARAPKMHVNGYVNRSLPGLIGCHRKSVYVHVGMTTPVHGFIFLEGHTFLVRN 299
Db 241 DPAPARAQAPAMHTVNGYVNRSLPGLIGCHKSVYVHVGMTSPVHGFIFLEGHTFLVRH 300
Qy 300 HQASLEISPTIFLTAOTLLMDLGOLFSLCHLSSHQHGMAYKVVDSCPRPQLRMKN 359
Db 301 HQASLEISPTIFLTAOTLLMDLGOLFSLCHLSSHHGMEAHVESCAEPQLRRKAD 360
Qy 360 EBAEDYDDDLTDSEMDVVRFDNDSFPTQIRSVAKKPKTWVHYIAAEERDWDYAPLVL 419
Db 361 EB-EDYDNLVSDMDVVRFDNDSFPTQIRSVAKKPKTWVHYIAAEERDWDYAPAVP 419
Qy 420 APDSRSYKQYVNGQRIKRYKVRMAYTDETFKTRAIQHSGLGILLYGEVGT 479
Db 420 SPDSRSYKSLVNSGQRIKRYKVRMAYTDETFKTRAIQHSGLGILLYGEVGT 479
Qy 480 LLIIFKNQASRPYNIYPHGITDVRPLYSRRLPKGVKHLKDPILPGLFKYKMTVVDG 539
Db 480 LLIIFKNQASRPYNIYPHGITDVSALHPCRLLKGVKHLKDPILPGLFKYKMTVVDG 539
Qy 540 PTKSDPRCLTRYSSYFVNMERDASGLIGPLLIICYKESVDQGNQIMSDKKNVILFSVD 599
Db 540 PTKSDPRCLTRYSSYFVNMERDASGLIGPLLIICYKESVDQGNQIMSDKKNVILFSVD 599

Qy	600	ENRSWYLFENIQRFLEPNPAGVCLDEPFOAGNIMESINGVYFDSLSLSVCLHEVAYWYL	659
Db	600	ENQSHWYLAENIQRFLEPNPDGQOPDEPFOAGNIMHSINGVYFDSLSLSVCLHEVAYWYL	659
Qy	660	SIGAQTDFLSVFFSGYTFKHVMVYEDTTLTLPFFSEIVFMSWENPGLWILGCHNSDRNR	719
Db	660	SVGAQTDFLSVFFSGYTFKHVMYEDTTLTLPFFSEIVFMSWENPGLWILGCHNSDLNR	719
Qy	720	GNTALLKVSSCDKNTGDIYEDSYEDISAYLLSXNNAIBPRFSQNSRHPSTROQFNATT	779
Db	720	GNTALLKVYSCDRDICDYDNTYEDIPGFLLSGKNVIEPRFQAQNSRPPSASQFOFII	779
Qy	780	IPENDIEXTDWFARHTBMPKIQNVSSDDLMLLROSTPHGLSLSDLOEAKYETFDSDP	839
Db	780	SPEDDVE-LDPOSGERTQALFELVPSGDSGNLGLQNPAPHGSSSSDLOEARN--ADDY	836
Qy	840	SFGAIDSNKSISEMTHFRPQLHHSQDMVFTPEGLQLRLNEKLTGTAATELKLKIDFKVSS	899
Db	837	LPGARERNTPASAAARLPELHSAERVLTPEP-----EK-----ELKLDKSWSS	882
Qy	900	TSNNLIS--TIPSDNLAAGTNTCSLGPSPMVHYDSOLDTTLTGKXSSPITBEGGPLSL	957
Db	883	SSDLTKSTPIPSDTLSAETRTHSLGQHPQVNFRSOLCAIVLGKSSHFIGAGVPLGS	942
Qy	958	SEENNDSKLLESGLMNSQSSWGNVSTBEGRLFKGKRAHPALLPKDNALFKVYSILL	1017
Db	943	TEEDF-----ESSLGENVSPVEDGIFEXERAHGPNASLTFRDDVLFKNVLSV	989
Qy	1018	KTKNTSNGATNRKTHIDGSPLLIENSPSVQNTILESDFEKVTPLIHDMMLDKNATA	1077
Db	990	KTKARVYLKTNRIHIDDAALLTENRASA-----TFWDKNTTA	1028
Qy	1078	LRANMHMKNTSSKNOMEWQCKGPIPPPAQNDPMDFKMLFLPESARWIORTHGKNSL	1137
Db	1029	SGLNHVSN-----WIKPLGKNPL	1047
Qy	1136	NSGQGPFPKQIVSLGPBKSVEGQNFLEKXNVVVGKEFTKDVGLKBMWVPSERNPLTN	1197
Db	1048	SSERGPPELLTSSGSKSVKGQSSGQGRIRVAVEBELSKG---KEMMLPNSLFTLTN	1104
Qy	1198	LDNIHENHTNQEKIOBRIEKKETLLIOENVLPQIHVTGTRKFMKNLFLLSRONVEG	1257
Db	1105	SADYQNGDTHSQGKSREMERREKLVQEKVDLPQVYTAGTKNFLENIHQSTEPSVEG	1164
Qy	1258	SYGAYAPVLQDPSLNDSTNRKTKTHAHSKKEEENLEGLNQTKQIWEKYACTTRIS	1317
Db	1165	FDGSHAPVQDSLNDSTNRAETHAHSJAREAPLEAGNRI-----	1210
Qy	1318	PNTSQNFVQTSRKALKQFRLPLEETELEKRIIVDDTSTQWSKMKHLTPSTLTQIDYN	1377
Db	1211	-GPGPRSAVPRRVKQSLKQIRLPLEBKIPERGVLNATSTRWS-----	1252
Qy	1378	EKEKGAITQSPLEDCLTRSHISIQANRSLPIAKVSSFPQIRPIYTRVLFQDNSSHLPA	1437
Db	1253	-----	1252
Qy	1438	ASYRKKDSGYQESSHFLOGAKKNLSALITLLEMTGDQREVGSLGTSATNSVTKKVENT	1497
Db	1253	-----ESSPILOGAKKNLSLPIFTLLEWAGQGISALGSAAGPLASGLEKA	1301
Qy	1498	VLPKPLDKTSGKVELLPKVHIYQKOLFFPDTETSGSPGHLDLVEGSLLOQTEGAIKNWEA	1557
Db	1302	VLSAGLSEASGKAEFLPKVRVHREDLLPKQTSNVSCAHDJQCEILFLQKTRGFVNKNV	1361
Qy	1558	NRGKVPFLAVATESAKTSKLLDPLAMDNHYGTOIPKEWKSQBKSPEKTAFFKKDRT	1617
Db	1362	NRPG-----RTPSKLLGP-----EMPK-EWBSLEKSPKSTALRTKDI	1398
Qy	1618	-LSINACESNHATAAINEGONKEPIEIVTWAKOGRTBELCSQNPPLVKHOREHTRITLOS	1676
Db	1399	SUPLDASHESNHSIAAKNVEGQAEYTCRGAAMTKQGGPRLCAPKXPVFLERHORDISLPTFO	1458

Qy	1677	QDSEIDYDDTISVBMKKEDFDIYDENQSPRFQKKTREHYTAAVERLWDYGMSSPHV	1736
Dd	1459	EEDDQYDDIFSTETKGEDEFDIYGEDENQDPRSQKTRHYTAAVEDQLWDYGMSSPRA	1518
Qy	1737	LRNEAQSGSVPOFKVYVQEFQTQGSFTQPLAYRGELNEHLLGPGYIAAEVEDNIMWTFRN	1796
Dd	1519	LRNEAQNGEVRFAKVYVREFADGFTQPSYRGELNKHLLGPGYIAAEVEDNIMWTFRN	1578
Qy	1797	QASRPYSFYSLIASEEDQDQGAEPKRNFKYPNETKTYFWKVQHHMAPTDEDFDCKAWAY	1856
Dd	1579	QASRPYSFYSLIAYPDQDQGAEPHNFVQPNETRYFWKVQHHMAPTDEDFDCKAWAY	1638
Qy	1857	PSDVLDEKDVHSGUIGPLNCHINTLNPAKGRQVTVQEFALPTTIEDTKSWTFTNNMER	1916
Dd	1639	PSDVLDEKDVHSGUIGPLLCRANTLNAAHGRQVTVQEFALPTTIEDTKSWTFTNNVER	1698
Qy	1917	NCRAPCNIQMEDPTFKENYRFAHNGIYIMDTLPGLVMAQDQRIWYLLMSGNSNENIHSIH	1976
Dd	1699	NCRAPCHLOMEDPTLKENYRFAHNGVYMDTLPGLVMAQNQRIWYLLMSGNSNENIHSIH	1758
Qy	1977	FSGHVFTVRKKEFKMALYNLYFGVETVEMLPKSKAGIWEVCLIEGHLHAGMSTLFLVY	2036
Dd	1759	FSGHVFTVRKKEFKMAVYNLYFGVETVEMLPKSKVIGWIECLI GELHQAQMSTTFLVY	1818
Qy	2037	SNKQOTPLGNASGHIRDFOITASQYQOWAPKLARLHYSGINAWSTKBPFSIKVDLLA	2096
Dd	1819	SKECAQPLGNASGRIRDFOITASQYQOWAPKLARLHYSGINAWSTKBPFSIKVDLLA	1878
Qy	2097	PMIIGHIKTOGARQKPSLSYISOFIIMYSLDGKKWQTYRGNSTGTLMVFPNGVNDSSGIKH	2156
Dd	1879	PMIIGHIMTOGARQKPSLSYISOFIIMYSLDGNWQSYRGNSTGTLMVFPNGVNDASGIKH	1938
Qy	2157	NINFPPIIARVIRLPHTHYSIRGTLRWELMGCDDLNSCMPLGWESKAISDAQITASSYFT	2216
Dd	1939	NINFPPIVABVIRLPHTHYSIRGTLRWELMGCDDLNSCMPLGONKAIISDQITASSHUS	1998
Qy	2217	NMFATWSPSKARLHULGGRSNAMPQVNNPKMIQVDFQKTMKVGTGTTQGVKSLTSMTV	2276
Dd	1999	NI FATWSPSQARLHULGGRSNAMPQVNSSAEMQLVDLQNTKVGTGTTQGVKSLTSSMYV	2058
Qy	2277	KEFLISSSDGHWTLFFONGKVYKFCQGNDSFTPTPVNSLDPELTRYLRIHPQSWHOI	2336
Dd	2059	KEFLISSSQGRWTLFLQDGHTKVFCQGNDSFTPVNALDPPLFTRYLRIEPTSWAQHI	2118
Qy	2337	ALRMEVLGCEAQDLY	2351
Dd	2119	ALRLEVIGCEAQDLY	2133

Search completed: April 13, 2004, 14:19:52
Job time : 73 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 13, 2004, 14:01:27 ; Search time 68.5 Seconds
(without alignments)
9697.360 Million cell updates/sec

Title: NP000123-328-355-581

Perfect score: 12415

Sequence: 1 MQIELSTCFPLLRFCFSA.....VVHQIALRMEVLGCEAQDLY 2351

Scoring table: BLOSUM62

Gapcp 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_29Jan04.*
1: geneseqp1980s.*
2: geneseqp1990s.*
3: geneseqp2000s.*
4: geneseqp2001s.*
5: geneseqp2002s.*
6: geneseqp2003as.*
7: geneseqp2003bs.*
8: geneseqp2004s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	12399	99.9	2351	1	AAP50059 Human fac
2	12399	99.9	2351	2	AAW11329 Native Pa
3	12399	99.9	2351	2	AAW46245 Human fac
4	12399	99.9	2351	2	AAW44373 Human fac
5	12399	99.9	2351	3	AAW52537 Human fac
6	12399	99.9	2351	6	AB995528 Amino aci
7	12395	99.8	2351	2	AAW11416 Active Pa
8	12395	99.8	2351	2	AAW11343 Active Pa
9	12395	99.8	2351	2	AAW11435 Active Pa
10	12394	99.8	2351	2	AAW11445 Active Pa
11	12394	99.8	2351	2	AAW11398 Active Pa
12	12394	99.8	2351	2	AAW11425 Active Pa
13	12394	99.8	2351	2	AAW11362 Active Pa
14	12394	99.8	2351	2	AAW11461 Active Pa
15	12394	99.8	2351	2	AAW11387 Active Pa
16	12394	99.8	2351	2	AAW11419 Active Pa
17	12393	99.8	2351	1	AAW81113 Factor VI
18	12393	99.8	2351	1	AAW80659 Sequence
19	12393	99.8	2351	2	AAW11408 Active Pa
20	12393	99.8	2351	2	AAW11427 Active Pa
21	12393	99.8	2351	2	AAW11347 Active Pa
22	12393	99.8	2351	2	AAW11332 Active Pa
23	12393	99.8	2351	2	AAW11437 Active Pa
24	12392	99.8	2351	2	AAW11371 Active Pa
25	12392	99.8	2351	2	AAW11396 Active Pa

ALIGNMENTS

RESULT 1

AAP50059
ID AAP50059 standard; protein; 2351 AA.

XX	AC	AAP50059;			
XX	XX				
DT	27-OCT-1991	(first entry)			
DE	Human factor VIII.				
XX	XX				
KW	Human factor VIII.				
XX	XX	Homc sapiens.			
XX	XX	EPI60457-A.			
XX	XX	06-NOV-1985.			
XX	XX	18-APR-1985;	85EP-00302734.		
PR	20-APR-1984;	84US-00602312.			
XX	XX	(GETH) GENENTECH INC.			
XX	XX	Capon DJ, Vehar GA, Lawn RM, Wood WI;			
XX	XX	WPI; 1985-277976/45.			
DR	N-P5DB; AAN50054.				
XX	XX	New recombinant functional human factor VIII or deriv. - useful for treating haemophilia and obtd. as pure prod. by recombinant DNA technology.			
PS	Disclosure; Fig 10a(10-10C(III)); 109pp; English.				
XX	XX	The sequence is that of human factor VIII. Amino acids 1-19 are the predicted signal peptide, and amino acids 1-2332 are the predicted mature protein. The protein is produced in pure form and in useful amts. using recombinant DNA techniques. Factor VIII can be used to correct factor VIII deficient plasma, and activates factor X to xa in the presence of factor IXa, Ca and phospholipid. These to activities are inactivated by antibodies specific for factor VIII. The activity of the prod. is bound to an immobilised monoclonal antibody specific for factor VIII. Factor VIII activity is activated by thrombin. The activity is bound to, and can be eluted from, immobilised von Willebrand factor. Dose of factor VIII is 20-40 units/kg over 8 hours i.v. for maintenance therapy for haemophilia, 40 units/kg for preoperative conditions, or 15-20 units/kg for minor haemorrhaging			

26	12390	99.8	2351	2	AAW00465 Factor-VI
27	12390	99.8	2351	2	AAW11454 Active Pa
28	12390	99.8	2351	4	AAW48843 Human fac
29	12390	99.8	2351	5	AAO18620 Human fac
30	12390	99.8	2351	6	ABR55853 Human fac
31	12389	99.8	2351	2	AAW11404 Active Pa
32	12389	99.8	2351	2	AAW11399 Active Pa
33	12389	99.8	2351	2	AAW11471 Active Pa
34	12388.5	99.8	2352	2	AAW11337 Active Pa
35	12388.5	99.8	2352	2	AAW11353 Active Pa
36	12388.5	99.8	2352	2	AAW11364 Active Pa
37	12388.5	99.8	2352	2	AAW11406 Active Pa
38	12388.5	99.8	2352	2	AAW11330 Active Pa
39	12388.5	99.8	2352	2	AAW11389 Active Pa
40	12388.5	99.8	2352	2	AAW11459 Active Pa
41	12388.5	99.8	2352	2	AAW11333 Active Pa
42	12388.5	99.8	2352	2	AAW11345 Active Pa
43	12388.5	99.8	2352	2	AAW11407 Active Pa
44	12388.5	99.8	2352	2	AAW11429 Active Pa
45	12388.5	99.8	2352	2	AAW11418 Active Pa

XX SQ Sequence 2351 AA;
 Query Match 99.9%; Score 12399; DB 1; Length 2351;
 Best Local Similarity 99.9%; Pred. No. 0;
 Matches 2348; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MQIELSTCFECLLRFCFSATRRYVLGAVELSDWYQSDLGELPVDARFPRVPKSPFFN 60
 DB 1 MQIELSTCFECLLRFCFSATRRYVLGAVELSDWYQSDLGELPVDARFPRVPKSPFFN 60

QY 61 TSVVYKTLFVETDHLFN--AKRPPPMWGLLPTQIAEVYDVTVITLKNWASHPVSLHAV 120
 DB 61 TSVVYKTLFVETDHLFN--AKRPPPMWGLLPTQIAEVYDVTVITLKNWASHPVSLHAV 120

QY 121 GSVYWKASAEAGVDDQTSOREKDDKVPFGSGSHYVQVLKENGPMASDPLCLTYSYLSH 190
 DB 121 GSVYWKASAEAGVDDQTSOREKDDKVPFGSGSHYVQVLKENGPMASDPLCLTYSYLSH 190

QY 181 VDLVKDLSGLIGALLVCREGSLAKEXTQTLHRKFIILLFAVFDGKSWHSETKNSLMQDRD 240
 DB 181 VDLVKDLSGLIGALLVCREGSLAKEXTQTLHRKFIILLFAVFDGKSWHSETKNSLMQDRD 240

QY 241 AASARAWPKMHTVNGYVNRSLPGLIGCHRSVYVHVIKGTTPVEHSIFLEGHTFLVRNH 300
 DB 241 AASARAWPKMHTVNGYVNRSLPGLIGCHRSVYVHVIKGTTPVEHSIFLEGHTFLVRNH 300

QY 301 RQASLEISPIITLTAQTLMDLQGLFLLSCHISSHQHDMGEAYVKVDSCPEPOLMKWNE 360
 DB 301 RQASLEISPIITLTAQTLMDLQGLFLLSCHISSHQHDMGEAYVKVDSCPEPOLMKWNE 360

QY 361 EADYDDDLTDSMDVVRFDNDSNPGFIQIRSVAKKPKTWVHYIAAEEDWDYAPLVLA 420
 DB 361 EADYDDDLTDSMDVVRFDNDSNPGFIQIRSVAKKPKTWVHYIAAEEDWDYAPLVLA 420

QY 421 PDERSYKSOVLNNGPQRIQKYYKAPFMAYDTDFPKTREA IQHESGILGPLLYGEGDIL 480
 DB 421 PDERSYKSOVLNNGPQRIQKYYKAPFMAYDTDFPKTREA IQHESGILGPLLYGEGDIL 480

QY 481 LIIFKQASRPYNIYPHGIDTVRPLPKGVKHLKOPFPLPGRIFKYKWTVVEDGP 540
 DB 481 LIIFKQASRPYNIYPHGIDTVRPLPKGVKHLKOPFPLPGRIFKYKWTVVEDGP 540

QY 541 TKSDPCLTRYSSFVNMREDLASGLIGPLLI CYKESVQKGNQIMSDKRNVLFSVFDE 600
 DB 541 TKSDPCLTRYSSFVNMREDLASGLIGPLLI CYKESVQKGNQIMSDKRNVLFSVFDE 600

QY 601 NRSWYLTENIQRELPNAGVOLBDPFOASNIMHSINGYVFDLSQLSVCLHEVAYWYILS 660
 DB 601 NRSWYLTENIQRELPNAGVOLBDPFOASNIMHSINGYVFDLSQLSVCLHEVAYWYILS 660

QY 661 IGAQTDPLSVFFSGYTFKHVMVYEDTLTLPFFSGETVFMSENPGLWILGCHNSDFRNG 720
 DB 661 IGAQTDPLSVFFSGYTFKHVMVYEDTLTLPFFSGETVFMSENPGLWILGCHNSDFRNG 720

QY 721 MYALLKVSXCDNTGYVSDSYEDI SAYLLSKNAIEPRFSFNSRHPSTROKQFNATTI 780
 DB 721 MYALLKVSXCDNTGYVSDSYEDI SAYLLSKNAIEPRFSFNSRHPSTROKQFNATTI 780

QY 781 PENDIEKTDPPFAHRTPMPKIQNVSSDDLMLLRQSPTEPHGLSLDQAKYETPSDDPS 840
 DB 781 PENDIEKTDPPFAHRTPMPKIQNVSSDDLMLLRQSPTEPHGLSLDQAKYETPSDDPS 840

QY 841 PGAIIDNNLSSEWTHFRPQLHSGDMVFTPEPSGLQLRLNEKLGTTAATLKKLDKFKVST 900
 DB 841 PGAIIDNNLSSEWTHFRPQLHSGDMVFTPEPSGLQLRLNEKLGTTAATLKKLDKFKVST 900

QY 901 SNNLJSTIPSDNLAAGTDNTSSLGPPSPMVHYDQSDTLTFLGKKSPLTESGGLSLEE 960
 DB 901 SNNLJSTIPSDNLAAGTDNTSSLGPPSPMVHYDQSDTLTFLGKKSPLTESGGLSLEE 960

QY 961 NNDKLLSGLMNSQBSWGNVSVSTESGRLFKGKAHGPALLTKDNALFKVYSISILKTN 1020
 DB 961 NNDKLLSGLMNSQBSWGNVSVSTESGRLFKGKAHGPALLTKDNALFKVYSISILKTN 1020

DB 961 NNDKLLSGLMNSQBSWGNVSVSTESGRLFKGKAHGPALLTKDNALFKVYSISILKTN 1020
 QY 1021 KTSNNSATNRKTHIDGPSLLIENSPSVWQNIILBSDTFFKKVTPLIHDRMLMDKNATALL 1080
 DB 1021 KTSNNSATNRKTHIDGPSLLIENSPSVWQNIILBSDTFFKKVTPLIHDRMLMDKNATALL 1080

QY 1081 NEMSNTKTSKKNEMVQOKKEGPIPPDAONPQMSPEKMLPLPESAWIORTGKNSLNG 1140
 DB 1081 NEMSNTKTSKKNEMVQOKKEGPIPPDAONPQMSPEKMLPLPESAWIORTGKNSLNG 1140

QY 1141 QPSPKQVLSLGEKSVQGNFLSEKNKVVGKGEFTKDVGLKEMVFPSPSRNLFLLNLDN 1200
 DB 1141 QPSPKQVLSLGEKSVQGNFLSEKNKVVGKGEFTKDVGLKEMVFPSPSRNLFLLNLDN 1200

QY 1201 LHENNTHNOEKIKQESIIEKKEKTLIQENVVLPOIHTVTGTFKNFMKNLFLSTEQNVGSD 1260
 DB 1201 LHENNTHNOEKIKQESIIEKKEKTLIQENVVLPOIHTVTGTFKNFMKNLFLSTEQNVGSD 1260

QY 1261 GAYAPVLODFRSLNDSTNFKKHTAHFSKKGSEENLEGLGNOTKOIVEKYACTRISPN 1320
 DB 1261 GAYAPVLODFRSLNDSTNFKKHTAHFSKKGSEENLEGLGNOTKOIVEKYACTRISPN 1320

QY 1321 SQONFVTOQSKKALKOFRPLBETELEKIIIVDDTSTQSKNMKHLTPSTLTQIDYNEKE 1380
 DB 1321 SQONFVTOQSKKALKOFRPLBETELEKIIIVDDTSTQSKNMKHLTPSTLTQIDYNEKE 1380

QY 1381 KGAIQTSPUSDCLTRSHSIPQANRSPPLPIAKVSSSPSTPIYLTAVLPDONSHPAAS 1440
 DB 1381 KGAIQTSPUSDCLTRSHSIPQANRSPPLPIAKVSSSPSTPIYLTAVLPDONSHPAAS 1440

QY 1441 RKKDSGVQBSHFLGAKKNNLSLAILTLEMTGDQREVGLSGLTSATNSVYKAKVENTVLP 1500
 DB 1441 RKKDSGVQBSHFLGAKKNNLSLAILTLEMTGDQREVGLSGLTSATNSVYKAKVENTVLP 1500

QY 1501 XPDLPKTSQKVELLPKVHIYQKDLPTTETNSGSPGHLDIVBGLSLQGTGAKWNEANRP 1560
 DB 1501 XPDLPKTSQKVELLPKVHIYQKDLPTTETNSGSPGHLDIVBGLSLQGTGAKWNEANRP 1560

QY 1561 GKVPFLRVATESAKTPSKLLDPLAWDNHYGQIPKEEKSKQEKPEKTAFAKKKTIISL 1620
 DB 1561 GKVPFLRVATESAKTPSKLLDPLAWDNHYGQIPKEEKSKQEKPEKTAFAKKKTIISL 1620

QY 1621 NACESNHAIAALINEGONKPEIEVTWAKQRTERLCSQNPVVKRQRETRITRITLOSORE 1680
 DB 1621 NACESNHAIAALINEGONKPEIEVTWAKQRTERLCSQNPVVKRQRETRITRITLOSORE 1680

QY 1681 IDYDDTISVEMKKEDPDYDEENQSPRSFQKTRHYTAAVERLWDYGMSSSPHVLNR 1740
 DB 1681 IDYDDTISVEMKKEDPDYDEENQSPRSFQKTRHYTAAVERLWDYGMSSSPHVLNR 1740

QY 1741 AQSGVPOPKVVOFDTGDSFTQPLYRGELNEHGLGPGYIRAEVEDNIMVTFEONOASR 1800
 DB 1741 AQSGVPOPKVVOFDTGDSFTQPLYRGELNEHGLGPGYIRAEVEDNIMVTFEONOASR 1800

QY 1801 PYSPYSSLIISYBEDORQGAEPKNTVKPNETKTYFWKVOHMAPTKDBPCKAWAYFSDV 1860
 DB 1801 PYSPYSSLIISYBEDORQGAEPKNTVKPNETKTYFWKVOHMAPTKDBPCKAWAYFSDV 1860

QY 1861 DLEKDVHSLGPLLVCHTNTLNPAGHQVTVQBPALFTTIDETKSWYFTENMERNCEA 1920
 DB 1861 DLEKDVHSLGPLLVCHTNTLNPAGHQVTVQBPALFTTIDETKSWYFTENMERNCEA 1920

QY 1921 PCNIQMEPTFKENYRFAHNGYIMDTLPGLVMAQDQIRWYLLSMGSENIHSHFSGH 1980
 DB 1921 PCNIQMEPTFKENYRFAHNGYIMDTLPGLVMAQDQIRWYLLSMGSENIHSHFSGH 1980

QY 1981 VFTVAKKXEYKVALNLYPGVFETVEMLPKAGIWRVECLIGELHLAGMSTLFLVYSNKC 2040
 DB 1981 VFTVAKKXEYKVALNLYPGVFETVEMLPKAGIWRVECLIGELHLAGMSTLFLVYSNKC 2040

QY 2041 QTPFLGWSGHIRDFOITASGQVQWAPKLARLHYSGSINAWSTKEPFSWKIVDLLAPMI 2100
 DB 2041 QTPFLGWSGHIRDFOITASGQVQWAPKLARLHYSGSINAWSTKEPFSWKIVDLLAPMI 2100

QY 2101 HGKIQGARKQESSLYISOFIIMYSLDGKKWOTYRGNSTGTLMVPPGNVDSSGIKHNIFN 2160
 DB 2101 HGKIQGARKQESSLYISOFIIMYSLDGKKWOTYRGNSTGTLMVPPGNVDSSGIKHNIFN 2160
 QY 2161 PPIIARYIRLHPHTYSIRSLRMELMGCDLNSCSMPLGWESKAISDAQITASSYFTNMYFA 2220
 DB 2161 PPIIARYIRLHPHTYSIRSLRMELMGCDLNSCSMPLGWESKAISDAQITASSYFTNMYFA 2220
 QY 2221 TWSPSKARLHLQGRNARFPQVNNPKWQLQVDFQKTMKVGTGTVGKSLTSMYVKFPL 2280
 DB 2221 TWSPSKARLHLQGRNARFPQVNNPKWQLQVDFQKTMKVGTGTVGKSLTSMYVKFPL 2280
 QY 2281 ISSSQDGHOWTLFFPONGKVKVPOGNQDSFTPVNSLDPPLLTRYLRIRHPQSMWHQIALRM 2340
 DB 2281 ISSSQDGHOWTLFFPONGKVKVPOGNQDSFTPVNSLDPPLLTRYLRIRHPQSMWHQIALRM 2340
 QY 2341 EVLGCEAODLY 2351
 DB 2341 EVLGCEAODLY 2351

RESULT 2

AAW11329
 ID AAW11329 standard; protein; 2351 AA.

AC AAW11329;

DT 17-NOV-1997 (first entry)

XX Native Factor VIII:C.

XX Factor VIII:C; analogue; glycoprotein; blood coagulation cascade;
 KW fibrinogen; fibrin clot; haemostasis; haemophilia A; bleeding diathesis;
 KW plasma protease; thrombin; immunogen; antibody; haemophilia; therapy;
 KW proteolytic cleavage.

OS Homo sapiens.

XX Key Location/Qualifiers

FT Peptide 1..19

FT /note= "signal peptide"

FT Protein 20..2351

FT /note= "mature Factor VIII:C"

FT Region 20..1667

FT /note= "heavy chain fragment"

FT Domain 760..1667

FT /note= "B domain"

FT Region 1668..2350

FT /note= "light chain fragment"

XX WO9703195-A1.

XX PN

XX 30-JAN-1997.

XX PD 09-JUL-1996; 96WC-US011444.

XX PF 11-JUL-1995; 95US-0001025P.

XX PR (CHIR) CHIRON CORP.

XX PA

XX Hung DT, Cohen FE, Innis M;

XX WIPI; 1997-119050/11.

XX DR N-PSDB; AAT51357.

XX Factor VIII:C analog modified adjacent to a non-activating Arg residue -

XX used in the treatment of haemophilias, by improvement of haemostasis.

XX Disclosure; Fig 1; 90pp; English.

XX This sequence represents the native Factor VIII:C. Factor VIII:C is a

XX large glycoprotein that participates in the blood coagulation cascade

CC that ultimately converts soluble fibrinogen to insoluble fibrin clot,
 CC effecting haemostasis. A deficiency in Factor VIII:C is responsible for
 CC haemophilia A, which is an X-chromosome-linked inherited bleeding
 CC diathesis. Factor VIII:C is activated by plasma proteases, such as
 CC thrombin. During activation the mature polypeptide is cleaved to generate
 CC heavy and light chain fragments that are further cleaved. The DNA
 CC encoding this sequence is mutated, using mutagenic primers, to produce
 CC the active Factor VIII:C analogues of the invention (such as AAW11330).
 CC The analogues comprise a native Factor VIII:C polypeptide modified at a
 CC site adjacent to a non-activating Arg residue so that a Arg-Pro or Pro-
 CC Arg dipeptide is created. Complexes of two or more of the analogues,
 CC nucleic acids and vectors encoding them may be used alone or in
 CC conjunction with each other, for the prevention or treatment of active
 CC Factor VIII:C deficiency in a mammal. The analogues may be used as
 CC immunogens to raise antibodies, and in the treatment of haemophilias, by
 CC improvement of haemostasis. The analogues are resistant to proteolytic
 CC cleavage and display increased plasma half-life. They may be administered
 CC at lower dosages and by different modes of administration
 XX
 XX Sequence 2351 AA;

Query Match 99.9%; Score 12399; DB 2; Length 2351;

Best Local Similarity 99.9%; Pred. No. 0;

Matches 2348; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MQIELSTCFELCLLRFCSATRRYYLGAVERISWDYMQSDLGELPVDARPPRPVKSPFPN 60
 DB 1 MQIELSTCFELCLLRFCSATRRYYLGAVERISWDYMQSDLGELPVDARPPRPVKSPFPN 60
 QY 61 TSVVYKKTLPVEFTDHLFENIAKPRPPNMGLLGPITQAEVYDVTVTILKNMASHPVSLHAV 120
 DB 61 TSVVYKKTLPVEFTDHLFENIAKPRPPNMGLLGPITQAEVYDVTVTILKNMASHPVSLHAV 120
 QY 121 GVSYYKASEGAYDDQTSQREKDDKVPGGSHYTVWQVLKENGSMASDPLCLTYSYLSH 180
 DB 121 GVSYYKASEGAYDDQTSQREKDDKVPGGSHYTVWQVLKENGSMASDPLCLTYSYLSH 180
 QY 181 VDLVKDNLNSLIGALLVCREGSLAKEKTQTLHKFILLFAVDPDEGKSHSETKNSLMQORD 240
 DB 181 VDLVKDNLNSLIGALLVCREGSLAKEKTQTLHKFILLFAVDPDEGKSHSETKNSLMQORD 240
 QY 241 AASARAWPKMHTVGYVNRSLPGLIGCHRKSVYVHVIWGMGTTPVHSIFLEGHTFLVRNH 300
 DB 241 AASARAWPKMHTVGYVNRSLPGLIGCHRKSVYVHVIWGMGTTPVHSIFLEGHTFLVRNH 300
 QY 301 ROASLEISPIITFLTAQTLMDLQGLLSCHTSSHQHDCWYAYKVVDSCPEEPQLKKNNE 360
 DB 301 ROASLEISPIITFLTAQTLMDLQGLLSCHTSSHQHDCWYAYKVVDSCPEEPQLKKNNE 360
 QY 361 EAEDYDDDLTDSXMDVVRFDNDNSPSFIQIRSVAKKHPKTVVHYIAAEEEDWDYAPLVLA 420
 DB 361 EAEDYDDDLTDSXMDVVRFDNDNSPSFIQIRSVAKKHPKTVVHYIAAEEEDWDYAPLVLA 420
 QY 421 PDRSYKQYLNNGPQRIQRYKKVRFMAYTDETFKTREAIQHSIGILGLLGEVGTDL 480
 DB 421 PDRSYKQYLNNGPQRIQRYKKVRFMAYTDETFKTREAIQHSIGILGLLGEVGTDL 480
 QY 481 LLIIFKNQASRPVNTYYPHGIITVRLPYSLRPLKGVKHLKDFPLPGELIFKYXWTVVREDGP 540
 DB 481 LLIIFKNQASRPVNTYYPHGIITVRLPYSLRPLKGVKHLKDFPLPGELIFKYXWTVVREDGP 540
 QY 541 TKSDPRCLTRYSSFPVNMERDLASGLIGPLLIQYKESVDQSGNQIMSDKRVNVLFSVFDE 600
 DB 541 TKSDPRCLTRYSSFPVNMERDLASGLIGPLLIQYKESVDQSGNQIMSDKRVNVLFSVFDE 600
 QY 601 NRSWYLTENIQRFIPNAGVQLEDPEFOASNMHMSINGYVFDLSQLSVCLHEVAYWYLS 660
 DB 601 NRSWYLTENIQRFIPNAGVQLEDPEFOASNMHMSINGYVFDLSQLSVCLHEVAYWYLS 660
 QY 661 IGAQTDLSLVSFSGYTFKHKVVYEDTLTPFSGETVFMSENCLWILGCHNSDFNRG 720
 DB 661 IGAQTDLSLVSFSGYTFKHKVVYEDTLTPFSGETVFMSENCLWILGCHNSDFNRG 720

QY 721 MTALLKVVSCDXTGTYEDSYEDISAYLLSKNNAIEPRFSQNSRHPSTROKQFNATTI 780
DB 721 MTALLKVVSCDXTGTYEDSYEDISAYLLSKNNAIEPRFSQNSRHPSTROKQFNATTI 780
QY 781 PENDIEKTDPAHRTMPKIQNVSSDILLMLLRQSPHPHGLSLDLQAKVETSDSDPS 840
DB 781 PENDIEKTDPAHRTMPKIQNVSSDILLMLLRQSPHPHGLSLDLQAKVETSDSDPS 840
QY 841 PGALDSNNLSSEMTFRPOLHSGDMVFTPESGIQLRLNEKLGTTAATLKKLDKPKVST 900
DB 841 PGALDSNNLSSEMTFRPOLHSGDMVFTPESGIQLRLNEKLGTTAATLKKLDKPKVST 900
QY 901 SNNLISTPSDNLAACTDNTSSLGPPSMPVHYDSQDITTLFGKSSPLTESGGPLSSEE 960
DB 901 SNNLISTPSDNLAACTDNTSSLGPPSMPVHYDSQDITTLFGKSSPLTESGGPLSSEE 960
QY 961 NNDKSLLEGLMNSQSSWGKNVSSTESGELFKGKAHGPALLTKDNALFKVSIISLLKTN 1020
DB 961 NNDKSLLEGLMNSQSSWGKNVSSTESGELFKGKAHGPALLTKDNALFKVSIISLLKTN 1020
QY 1021 KTSNNSATNRKTHIDGFSLLIENSPSVWQNI LESDTEFKKVTPLIHDRMLMDKNATALRL 1080
DB 1021 KTSNNSATNRKTHIDGFSLLIENSPSVWQNI LESDTEFKKVTPLIHDRMLMDKNATALRL 1080
QY 1081 NMSNKTTSKNMVMYQKKEGIPDDAQNPDMSFFKMLFLPESARWIQRTGKNSLNSG 1140
DB 1081 NMSNKTTSKNMVMYQKKEGIPDDAQNPDMSFFKMLFLPESARWIQRTGKNSLNSG 1140
QY 1141 QGSPKQVLSGPEKVEGQNFISEKNKVVVGGEETKQGLKEMVPPSSNIFLTNLN 1200
DB 1141 QGSPKQVLSGPEKVEGQNFISEKNKVVVGGEETKQGLKEMVPPSSNIFLTNLN 1200
QY 1201 LHENNTNHEKKIQEIEKKEKTELQENVLPQIHTVTGKTFMKNLFLASTRONVEGSD 1260
DB 1201 LHENNTNHEKKIQEIEKKEKTELQENVLPQIHTVTGKTFMKNLFLASTRONVEGSD 1260
QY 1261 GAYAPVLQDPRSLNDSTNRKTAHFSKKEBENLEGLNQTKQIIVEKYACTRISPNT 1320
DB 1261 GAYAPVLQDPRSLNDSTNRKTAHFSKKEBENLEGLNQTKQIIVEKYACTRISPNT 1320
QY 1321 SQQNFVQTSKRALKOPRLPLEBTELEKRIIVDDTSQMSKNNKELTPSLTQIDYNEKE 1380
DB 1321 SQQNFVQTSKRALKOPRLPLEBTELEKRIIVDDTSQMSKNNKELTPSLTQIDYNEKE 1380
QY 1381 KGAIQSPSLSDCLTRSHSIPQANRSLPIAKVSSFPSIRPIYITRVLFDNNSHLPAASY 1440
DB 1381 KGAIQSPSLSDCLTRSHSIPQANRSLPIAKVSSFPSIRPIYITRVLFDNNSHLPAASY 1440
QY 1441 RKDQSGVQESSHFLQAKAKNNLSLAILTLEMTGDQREVGLGTSATNSVTYKKVENTVLP 1500
DB 1441 RKDQSGVQESSHFLQAKAKNNLSLAILTLEMTGDQREVGLGTSATNSVTYKKVENTVLP 1500
QY 1501 KPDLPTSGKVELLPKVHIYOKDLFPPTETSGSPGHLDLVEGSLLOQTEGAIKWNEANRP 1560
DB 1501 KPDLPTSGKVELLPKVHIYOKDLFPPTETSGSPGHLDLVEGSLLOQTEGAIKWNEANRP 1560
QY 1561 GKVPFLVATESSAKTSPSKLLDPLAWDNHYGTQIPKEEMKSQKSPKTAFFKKDKTILSL 1620
DB 1561 GKVPFLVATESSAKTSPSKLLDPLAWDNHYGTQIPKEEMKSQKSPKTAFFKKDKTILSL 1620
QY 1621 NACESNHAIAINEGQNKPIEVTWAKQGRTERLCSQNPVLYKHQBREITRTLQSDQEE 1680
DB 1621 NACESNHAIAINEGQNKPIEVTWAKQGRTERLCSQNPVLYKHQBREITRTLQSDQEE 1680
QY 1681 IDYDDTISVEMKKEDFDIYEDENQSPRSFOKTRHYFIAAVERLMDYGMSSSPHVLNR 1740
DB 1681 IDYDDTISVEMKKEDFDIYEDENQSPRSFOKTRHYFIAAVERLMDYGMSSSPHVLNR 1740
QY 1741 AQSGSVPOFKKVVQFEFTDGSFTQPLYRGELNDELGLLGPYIRAEVEDNIVVTFRNQASR 1800
DB 1741 AQSGSVPOFKKVVQFEFTDGSFTQPLYRGELNDELGLLGPYIRAEVEDNIVVTFRNQASR 1800
QY 1801 PYSFYSSLISYEDQRCQAEPKNEKTYFWKQVHHMAPTKDFDCKAWAFSDV 1860

DB 1801 PYSFYSSLISYEDQRCQAEPKNEKTYFWKQVHHMAPTKDFDCKAWAFSDV 1860
QY 1861 DLEKDVHSLGLPILAVCHTNTINPAHGQVTVQEZALPFTTFDETKSWYFTFEMERNCR 1920
DB 1861 DLEKDVHSLGLPILAVCHTNTINPAHGQVTVQEZALPFTTFDETKSWYFTFEMERNCR 1920
QY 1921 PCNIQMEPTPKENYRFAINGYIMDTLPGVMAOQRIWYLLSMGNGENIHSIHFSGH 1980
DB 1921 PCNIQMEPTPKENYRFAINGYIMDTLPGVMAOQRIWYLLSMGNGENIHSIHFSGH 1980
QY 1981 VFTVRKXBEYKMALXNLYPGVPEVEMLPKAGIWRVECLIGEHLHAGNSTLFLVYSNKC 2040
DB 1981 VFTVRKXBEYKMALXNLYPGVPEVEMLPKAGIWRVECLIGEHLHAGNSTLFLVYSNKC 2040
QY 2041 QPPLGWASGHIRDFQITASGOYQOWAPKLARLHYSGSINAWSTKEPFSWIKVDLLAPMI 2100
DB 2041 QPPLGWASGHIRDFQITASGOYQOWAPKLARLHYSGSINAWSTKEPFSWIKVDLLAPMI 2100
QY 2101 HGIKTQGARQKFESSVISOFTIMVSLDGKKWQTYRGNSTGTLMVFFGNVDSSGIKENIFN 2160
DB 2101 HGIKTQGARQKFESSVISOFTIMVSLDGKKWQTYRGNSTGTLMVFFGNVDSSGIKENIFN 2160
QY 2161 PPIIARYIRLHPHTHYSIRSTLRMELMGCGLNCSMPLGWESKAI SDAQITASSYFTNMPA 2220
DB 2161 PPIIARYIRLHPHTHYSIRSTLRMELMGCGLNCSMPLGWESKAI SDAQITASSYFTNMPA 2220
QY 2221 TWSPSKARLHLQGRNANRPOVNNPEKMOYDFQKTMKVGTTCQVKSLLTSMYVKFEL 2280
DB 2221 TWSPSKARLHLQGRNANRPOVNNPEKMOYDFQKTMKVGTTCQVKSLLTSMYVKFEL 2280
QY 2281 ISSSQDGHQWTLFFQNGKVKVFGQNGQDFTPVVNSLDPLLLTRYLRIHPQSVWHQIALRM 2340
DB 2281 ISSSQDGHQWTLFFQNGKVKVFGQNGQDFTPVVNSLDPLLLTRYLRIHPQSVWHQIALRM 2340
QY 2341 EYLGCEAQDLY 2351
DB 2341 EYLGCEAQDLY 2351
RESULT 3
AAW46245
ID AAW46245 standard; protein; 2351 AA.
XX AAW46245;
AC XX
XX 25-MAR-2003 (revised)
DT 06-AUG-1998 (first entry)
XX XX
DE Human factor III protein sequence.
XX Replication defective; recombinant retrovirus; RRV; therapeutic protein;
KW haemophilia; thrombosis; hypercoagulable disorder; liver disease; human;
KW hepatitis; thalassemia; phenylketonuria; Lesch-Nyhan syndrome; diabetes;
KW cystic fibrosis; Duchenne's Muscular Dystrophy; hypercholesterolemia;
KW hypopituitarism; adenine deaminase deficiency; HIV infection; anaemia;
KW Guacher's syndrome; high blood pressure; Alzheimer's disease, factor III;
KW autoimmune; inflammatory disease.
XX Homo sapiens.
XX XX
FN WO9800541-A2.
XX XX
PD 08-JAN-1998.
XX XX
PF 02-JUL-1997; 97WC-US011784.
XX XX
PR 03-JUL-1996; 96DS-00645601.
PR 13-AUG-1996; 96US-00696381.
PR 04-JUN-1997; 97US-00369309.
XX XX
PA (CHIR) CHIRON CORP.
XX XX

PI Jolly DU, Barber JR, Chang SMW, Respass JG, Allen JR, Boder M;
PI Chong K, De La Vega D, Depolo NJ, Hsu DC, Ibanes CE;
PI Mittelstaedt DM, Prussak CE, Greengard J, Lee R;
XX
DR WPI: 1998-086966/08.
DR N-PSDB; AAV19580.

XX New replication defective recombinant retro-viruses - which can be
XX administered to provide long term systemic expression of therapeutic
XX protein in blood, useful in, e.g. treating hyper-coagulable disorders.
XX
XX Example 28; Page 203-210; 272pp; English.

XX This is the human factor III sequence. The encoding DNA is used in the
XX construction of recombinant retroviral vectors expressing human factor
XX VIII. The specification provides the preparation of replication defective
XX recombinant retrovirus (RRV) expressing a therapeutic protein. The RRV
XX preparation is resistant to degradation by human complement and is
XX capable of inducing long term systemic expression of the therapeutic
XX protein when administered intravenously to a human. The long term
XX systemic expression results in a measurable level of the therapeutic
XX protein being produced in the blood of the human for a period of at least
XX 30 days after the administration of the RRV vector preparation. RRV's can
XX be used for in vivo delivery of therapeutic protein to treat, e.g.
XX haemophilia A, haemophilia B, thrombosis, hypercoagulable disorders,
XX liver diseases such as hepatitis, disorders such as thalassemia,
XX phenylketonuria, Lesch-Nyhan syndrome, severe combined immunodeficiency
XX (SCID), cystic fibrosis, Duchenne's Muscular Dystrophy, inherited
XX emphysema, familial hypercholesterolemia, diabetes, hypopituitarism,
XX adenine deaminase deficiency, algal-antitrypsin deficiency, Gaucher's
XX syndrome, anaemia, infections such as HIV infection, high blood pressure,
XX Alzheimer's disease, autoimmune or inflammatory disease or graft versus
XX host disease. RRV's are capable of surviving inactivation in human serum
XX thereby allowing efficient gene transfer over prolonged periods of time.
XX (Updated on 25-MAR-2003 to correct PI field.)

XX Sequence 2351 AA;

Query Match 99.9%; Score 12399; DB 2; Length 2351;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 2348; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MQIELSTCFFCLLRFCFSATRRYVIGAVELSWYMQSDGLGELPVDARFPVRVPSFPPN 60
DB 1 MQIELSTCFFCLLRFCFSATRRYVIGAVELSWYMQSDGLGELPVDARFPVRVPSFPPN 60

QY 61 TSVVYKTLFVEFTDHLNIAKPRPPWMLGPTTQAEVDTWVTLKKNMASHPVSLHAV 120
DB 61 TSVVYKTLFVEFTDHLNIAKPRPPWMLGPTTQAEVDTWVTLKKNMASHPVSLHAV 120

QY 121 GVSYYKASGEAYDDQTSOREKEDDKVPFGGSHTYVWVLYKENGWASDPLCLTYSYLSH 180
DB 121 GVSYYKASGEAYDDQTSOREKEDDKVPFGGSHTYVWVLYKENGWASDPLCLTYSYLSH 180

QY 181 VDLVKDINSGLIGALLVCREGLAKEKTQTLHKFILLFAVDEGKSWHSETKNSLMQDRD 240
DB 181 VDLVKDINSGLIGALLVCREGLAKEKTQTLHKFILLFAVDEGKSWHSETKNSLMQDRD 240

QY 241 AASARAWPKHTVNGVYVNRSLPGLIGCHRKSVYWHVIGMTTPVHSIFLEGHTFLVRNH 300
DB 241 AASARAWPKHTVNGVYVNRSLPGLIGCHRKSVYWHVIGMTTPVHSIFLEGHTFLVRNH 300

QY 301 RQASLEISPIITLTAQTLLMDLQFLLSCHISSHQHDMGEAVVKYVDSQCEEPQIMKXNE 360
DB 301 RQASLEISPIITLTAQTLLMDLQFLLSCHISSHQHDMGEAVVKYVDSQCEEPQIMKXNE 360

QY 361 EAEDYDDDLTDEMDVVRPDDNNSPSFQIRSVAKKPKTWVHYTAABEEDWDYAPLVIA 420
DB 361 EAEDYDDDLTDEMDVVRPDDNNSPSFQIRSVAKKPKTWVHYTAABEEDWDYAPLVIA 420

QY 421 PDRSVKQYLNNGPQRCGRKKYKVRMAVDTETFKTRAQCHESGILGPLLYGVBGDTL 480
DB 421 PDRSVKQYLNNGPQRCGRKKYKVRMAVDTETFKTRAQCHESGILGPLLYGVBGDTL 480

QY 481 LIIFKNQASRPYNIYPHGITDVRPLYSRRLPKGVXHLKDPFPLPGRIEIKYKWTVTVEDGP 540
DB 481 LIIFKNQASRPYNIYPHGITDVRPLYSRRLPKGVXHLKDPFPLPGRIEIKYKWTVTVEDGP 540

QY 541 TKSDPRCLTRYVSSFVNMRDLASGLIGPLLCYKESVDCKGNQJMSDKRNVILSVFDE 600
DB 541 TKSDPRCLTRYVSSFVNMRDLASGLIGPLLCYKESVDCKGNQJMSDKRNVILSVFDE 600

QY 601 NRSWYLTENIQRFLEPNPAGVQLEDPEFQASNMHSINGVYVDSLSQSLVCLHEVAVYIIS 660
DB 601 NRSWYLTENIQRFLEPNPAGVQLEDPEFQASNMHSINGVYVDSLSQSLVCLHEVAVYIIS 660

QY 661 IGAQTDPLSVFSGVTFKHKWYEDTILFPFSGHTVFMENPGLWILGCHNSFRNRG 720
DB 661 IGAQTDPLSVFSGVTFKHKWYEDTILFPFSGHTVFMENPGLWILGCHNSFRNRG 720

QY 721 MTALLKVSSCDKNTGDDYEDSVEDISAYLLSKNNAIEPRSPNSRHPSTKQKNATTI 780
DB 721 MTALLKVSSCDKNTGDDYEDSVEDISAYLLSKNNAIEPRSPNSRHPSTKQKNATTI 780

QY 781 PENDIEKTDMPFAHRTMPKTONVSSDMLLRQSPTPHGLSLSDLOEAKYETFSDDPS 840
DB 781 PENDIEKTDMPFAHRTMPKTONVSSDMLLRQSPTPHGLSLSDLOEAKYETFSDDPS 840

QY 841 PGADSNNSLSEMTFRPOLHSCDMVTPESGLQRLNEKLGTTAAATELKLPKVSST 900
DB 841 PGADSNNSLSEMTFRPOLHSCDMVTPESGLQRLNEKLGTTAAATELKLPKVSST 900

QY 901 SNNLSTIPSNLAAGTNTSSLPSPMPHYVDSQDPTTLFGKSSSPITESGGPLSLEE 960
DB 901 SNNLSTIPSNLAAGTNTSSLPSPMPHYVDSQDPTTLFGKSSSPITESGGPLSLEE 960

QY 961 NNDKLLBSGLMNSQESSGKNVSTESGRLPKGRAGHPALLTKONALFKVSISSLATN 1020
DB 961 NNDKLLBSGLMNSQESSGKNVSTESGRLPKGRAGHPALLTKONALFKVSISSLATN 1020

QY 1021 KTSNNSATNRKTHIDGSPSLJJTENSPPVWQNTLESDETEKKVTPLIHDRLMDKNAATRL 1080
DB 1021 KTSNNSATNRKTHIDGSPSLJJTENSPPVWQNTLESDETEKKVTPLIHDRLMDKNAATRL 1080

QY 1081 NEMSNKTTSSKNMVMVQCKEGPIPPDAQNPDMPFFKMLFLPESARWQRTGKNSLNSG 1140
DB 1081 NEMSNKTTSSKNMVMVQCKEGPIPPDAQNPDMPFFKMLFLPESARWQRTGKNSLNSG 1140

QY 1141 QGPFKQLVSLGPKSVGQNFLEKKNKVVGKGEFTKDVGLKENVFPSSNNLFLTNLN 1200
DB 1141 QGPFKQLVSLGPKSVGQNFLEKKNKVVGKGEFTKDVGLKENVFPSSNNLFLTNLN 1200

QY 1201 LHENNTNQEKKIOEEIEKKEITLQENVVLPQIHTVTGTKNPMKNLFLSTRONVEGSVD 1260
DB 1201 LHENNTNQEKKIOEEIEKKEITLQENVVLPQIHTVTGTKNPMKNLFLSTRONVEGSVD 1260

QY 1261 GAYAPVLQDPFSLANDSTNRKTKHTAHFSPKSGEENLEGNGQTKQIVEKYACTTRISENT 1320
DB 1261 GAYAPVLQDPFSLANDSTNRKTKHTAHFSPKSGEENLEGNGQTKQIVEKYACTTRISENT 1320

QY 1321 SQOQFVTVORSKRALKQFELPLETELEKRIITVDSTSTQWSKNMKHLTFTSTLTQIDYNEKE 1380
DB 1321 SQOQFVTVORSKRALKQFELPLETELEKRIITVDSTSTQWSKNMKHLTFTSTLTQIDYNEKE 1380

QY 1381 KGATITQSPLSCLTRSHSIPOANRSPPIAKVSSFPSPRIPIYLRVLFDQNSHLPAAZY 1440
DB 1381 KGATITQSPLSCLTRSHSIPOANRSPPIAKVSSFPSPRIPIYLRVLFDQNSHLPAAZY 1440

QY 1441 RKXDSGVQESSHFIQGAKKNNLSLAITLMTGQREYVGLSGTSATNSVTYKVKVENTVLP 1500
DB 1441 RKXDSGVQESSHFIQGAKKNNLSLAITLMTGQREYVGLSGTSATNSVTYKVKVENTVLP 1500

QY 1501 KPDLFKTSKGKVELLPKVHIYOKDLFPPTETSNVSGSPGHLDLVEGSLQGTGEGAIKWNEANRP 1560
DB 1501 KPDLFKTSKGKVELLPKVHIYOKDLFPPTETSNVSGSPGHLDLVEGSLQGTGEGAIKWNEANRP 1560


```
QY 1561 GKVPFLRVATESSAKTPSKLLDPLANDNHYGTQIPKEWASQKSPKTAFAKKKOTILSL 1620
Db 1561 GKVPFLRVATESSAKTPSKLLDPLANDNHYGTQIPKEWASQKSPKTAFAKKKOTILSL 1620
QY 1621 NACESNHAI AAINEGQNKPIEYTWAKQTERHLCSONPPVLKQHOREIRTRTLQSDQEE 1680
Db 1621 NACESNHAI AAINEGQNKPIEYTWAKQTERHLCSONPPVLKQHOREIRTRTLQSDQEE 1680
QY 1681 IDYDDT-SVEMKEDDFDIYEDENQSPRSQKTRHYFIAAVERLDYGMSSSPHVLNR 1740
Db 1681 IDYDDT-SVEMKEDDFDIYEDENQSPRSQKTRHYFIAAVERLDYGMSSSPHVLNR 1740
QY 1741 AQSGVPQFKKVVQFQFTDGSFTQPLVRGELNHLGLLGFYIRAEVEDNIMVTRNQASR 1800
Db 1741 AQSGVPQFKKVVQFQFTDGSFTQPLVRGELNHLGLLGFYIRAEVEDNIMVTRNQASR 1800
QY 1801 PYSFYSLSIYVEDORQGAEPKRNFKYKNTKTYFKVQVQHMAPTKDEDFCKAWAYFSDV 1860
Db 1801 PYSFYSLSIYBEDQGAEPKRNFKYKNTKTYFKVQVQHMAPTKDEDFCKAWAYFSDV 1860
QY 1861 DLEKDVHSGLIGLPLLVCHTNTLPAHGRQVTVQEFALFETIPDETWSYFTEVNEENCR 1920
Db 1861 DLEKDVHSGLIGLPLLVCHTNTLPAHGRQVTVQEFALFETIPDETWSYFTEVNEENCR 1920
QY 1921 PCNIQMEDPTFKENYRFAHNGYIMDTPLGLVVAQDQIRWYLLSMGSENIHISIFSGH 1980
Db 1921 PCNIQMEDPTFKENYRFAHNGYIMDTPLGLVVAQDQIRWYLLSMGSENIHISIFSGH 1980
QY 1981 VFTVRKKEEYKALYNLYPGVFETVEMLPKAGIWRVECLIGHLHAGHSTLFLVTSNKC 2040
Db 1981 VFTVRKKEEYKALYNLYPGVFETVEMLPKAGIWRVECLIGHLHAGHSTLFLVTSNKC 2040
QY 2041 QTPGLMASGHIRDFQITASQYQWAPKPLARLHYSGSINAWSTKEPFSWIKVDLLAPMII 2100
Db 2041 QTPGLMASGHIRDFQITASQYQWAPKPLARLHYSGSINAWSTKEPFSWIKVDLLAPMII 2100
QY 2101 HGKTCQAKQKFSLSIYISQPIIMYSLDGKKWQYRGNSTGTMVFFGNVDSSGIKENIFN 2160
Db 2101 HGKTCQAKQKFSLSIYISQPIIMYSLDGKKWQYRGNSTGTMVFFGNVDSSGIKENIFN 2160
QY 2161 PPIIARYIRLHPTHYSIRSLRMLGCDLNSCMLPGMESKAIQITASSYFTNMPA 2220
Db 2161 PPIIARYIRLHPTHYSIRSLRMLGCDLNSCMLPGMESKAIQITASSYFTNMPA 2220
QY 2221 TWSPSKARLHLOGRSNARPVQNNKQWLDQVDPQKTMKVTGVTQGVKSLITSYVYK3FL 2280
Db 2221 TWSPSKARLHLOGRSNARPVQNNKQWLDQVDPQKTMKVTGVTQGVKSLITSYVYK3FL 2280
QY 2281 ISSQDGHQWTLFPQNGKVKVFGNQDSFTPVVNSLDPPLLTRYLRIHPQSWVHQIALRM 2340
Db 2281 ISSQDGHQWTLFPQNGKVKVFGNQDSFTPVVNSLDPPLLTRYLRIHPQSWVHQIALRM 2340
QY 2341 EVLGCEAQDLY 2351
Db 2341 EVLGCEAQDLY 2351
```

RESULT 4

AAW44373

ID AAW44373 standard; protein; 2351 AA.

XX AC AAW44373;

XX DT 20-JUL-1998 (first entry)

XX DE Human Factor VIII.

XX KW Factor VIII; blood clotting; haemophilia A; gene therapy; retrovirus;

XX KW vector; human.

XX OS Homo sapiens.

XX FH Key

Location/Qualifiers

```
Domain 711..1694
/Note= "B-domain"
PN W09800542-A2.
XX 08-JAN-1998.
XX 02-JUL-1997; 97NO-US011785.
XX 03-JUL-1996; 96US-00545601.
PR 13-AUG-1996; 96US-00596381.
PR 04-JUN-1997; 97US-00369309.
XX (CHIR ) CHIRON CORP.
XX Jolly DJ, Barber JR, Chang SM, Respass JG, Allen JR, Boder M;
PI Chong K, De La Vega D, Depolo NJ, Hsu DC, Ibanez CE;
PI Mittelstaedt DM, Prussak CE, Greengard J;
XX WPI: 1998-086967/08.
DR N-PSDB; AAV15359.
XX New replication defective recombinant retroviruses - which express B
domain-deleted human factor VIII or human factor IX for the treatment of
haemophilia.
PS Example 28; Page 164-166; 236pp; English.
XX This polypeptide comprises human Factor VIII. The invention relates to
CC preparations of replication defective recombinant retrovirus (RV)
CC expressing a B domain-deleted human Factor VIII protein, such as the SQN
CC deletion mutant (see AAM44372), where the recombinant RV is capable of
CC infecting human cells, is resistant to degradation by human complement
CC and is capable of inducing long-term (at least 30 days and up to 6 months
CC or longer post-injection) systemic expression of Factor VIII when
CC administered to a haemophilia A patient
XX Sequence 2351 AA;
Query Match 99.9%; Score 12399; DB 2; Length 2351;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 2348; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
QY 1 MOELSTCFCLLAFCSATRRYYLGAVELSWDMQSDGLBPLVDARPPRPVKSFPFN 60
Db 1 MOELSTCFCLLAFCSATRRYYLGAVELSWDMQSDGLBPLVDARPPRPVKSFPFN 60
QY 61 TSVVYKTLFVEFTDELFNIAKPRPPWMLGLOPTTQAEVYDTVITLKNMASHPVSLHAV 120
Db 61 TSVVYKTLFVEFTDELFNIAKPRPPWMLGLOPTTQAEVYDTVITLKNMASHPVSLHAV 120
QY 121 GVSVMKASGAEYDDQTSQREKDDKVPFGSGHTYVQVLKENGMSADPLCLTYSYLSH 180
Db 121 GVSVMKASGAEYDDQTSQREKDDKVPFGSGHTYVQVLKENGMSADPLCLTYSYLSH 180
QY 181 VDLVKDLSGLIGALLVCREGSLAKEKTQTLHKFILLFAVDEGKSWSETKNSLMQDRD 240
Db 181 VDLVKDLSGLIGALLVCREGSLAKEKTQTLHKFILLFAVDEGKSWSETKNSLMQDRD 240
QY 241 AASAAWPKMHTVNGYVNRSLPGLIGCHRKSVYWHVIGMTTPEVHSIFLEGHTFLVRNH 300
Db 241 AASAAWPKMHTVNGYVNRSLPGLIGCHRKSVYWHVIGMTTPEVHSIFLEGHTFLVRNH 300
QY 301 ROASLEISPIITELCACTLLMDLQFLLSCHTSSHQHGDGMEAVYKVDSCPEEPOLIMKNE 360
Db 301 ROASLEISPIITELCACTLLMDLQFLLSCHTSSHQHGDGMEAVYKVDSCPEEPOLIMKNE 360
QY 361 EAEDYDDDLTDSMDVVRPDDNDSFQIRSVAKKHPTKWVHYIAAEEEDWDYAPLVIA 420
Db 361 EAEDYDDDLTDSMDVVRPDDNDSFQIRSVAKKHPTKWVHYIAAEEEDWDYAPLVIA 420
QY 421 PDRSYKSOYLNGPQIRGRKYKVRFMAYTDETPKTEAIOHRSGLIGPLLYGBVGTDL 480
Db 421 PDRSYKSOYLNGPQIRGRKYKVRFMAYTDETPKTEAIOHRSGLIGPLLYGBVGTDL 480
```

Db 421 PDRSKYKQYLNQGRIGRKYKVRPMAYTDEFTKTRAIQHESGILGPELLYGEVGDTL 480
QY 481 LIIFKQASRPYNIYEHGIDVRPLSRRLPKGVKHLKDPILPGETFKYKWTVTVEDGP 540
Db 481 LIIFKQASRPYNIYEHGIDVRPLSRRLPKGVKHLKDPILPGETFKYKWTVTVEDGP 540
QY 541 TKSDPRCLTRYSSFVNMRDLASGLIGPLLI CYKESVDQGNQIXSDKENVILFSVFDE 600
Db 541 TKSDPRCLTRYSSFVNMRDLASGLIGPLLI CYKESVDQGNQIXSDKENVILFSVFDE 600
QY 601 NRSWYLTENIORLNPAGVQLEDDPFOASNIHMSINGYVFDLSQISVCLHEVAYWYLS 660
Db 601 NRSWYLTENIORLNPAGVQLEDDPFOASNIHMSINGYVFDLSQISVCLHEVAYWYLS 660
QY 661 IGAQTDFLSVFFSGYTFKHKWYEDTLTLPFGGETVFMGMENPGLWILGCHNSDPRNG 720
Db 661 IGAQTDFLSVFFSGYTFKHKWYEDTLTLPFGGETVFMGMENPGLWILGCHNSDPRNG 720
QY 721 MTALLKVSXCDKNTGYEDSYEDISAYLLSKNAIEPRSFQNSRHPSTQKQFNATTI 780
Db 721 MTALLKVSXCDKNTGYEDSYEDISAYLLSKNAIEPRSFQNSRHPSTQKQFNATTI 780
QY 781 PENDIEKTDWFAHRTMPMKIQNVSSDILLMLRQSTPHCLSLSDLOBAXYTFSDDP 840
Db 781 PENDIEKTDWFAHRTMPMKIQNVSSDILLMLRQSTPHCLSLSDLOBAXYTFSDDP 840
QY 841 PGADSNNSLSEMTFRPOLHSGDMVFTPESGQLRLNEKIGTAAATELKKLDFKVSST 900
Db 841 PGADSNNSLSEMTFRPOLHSGDMVFTPESGQLRLNEKIGTAAATELKKLDFKVSST 900
QY 901 SNNLSTIFSDNLAAGTDNYSIGPSPMPVHYDSQDITLFGKKSPLTESGGLFSLEE 960
Db 901 SNNLSTIFSDNLAAGTDNYSIGPSPMPVHYDSQDITLFGKKSPLTESGGLFSLEE 960
QY 961 NNDSKLLLEGGLMNSQSSGKNVSSSTESGRLPKGAHGPALLTKNALFKVISILKTN 1020
Db 961 NNDSKLLLEGGLMNSQSSGKNVSSSTESGRLPKGAHGPALLTKNALFKVISILKTN 1020
QY 1021 XTNNSATNRKTHIDGPSLLIENSPPVWQNILESDETFPKKVTPLIHDRMLMDKNAALRL 1080
Db 1021 XTNNSATNRKTHIDGPSLLIENSPPVWQNILESDETFPKKVTPLIHDRMLMDKNAALRL 1080
QY 1081 NKSANKTSSKNMEMYQOKKEGIPDPAQNDMSFFKMLFPESAAMIORTHGKNSLNSG 1140
Db 1081 NKSANKTSSKNMEMYQOKKEGIPDPAQNDMSFFKMLFPESAAMIORTHGKNSLNSG 1140
QY 1141 QGSPKQVSLGPEKVEGQNFLENKVVVGKGEFTKDVGLKEMVFPSSRLFLTNDN 1200
Db 1141 QGSPKQVSLGPEKVEGQNFLENKVVVGKGEFTKDVGLKEMVFPSSRLFLTNDN 1200
QY 1201 LHNNTHNQEKTIQBEIRKKEKTLIQENVLPQIHTVTGKTFMKNLFLASTRQNVGSYD 1260
Db 1201 LHNNTHNQEKTIQBEIRKKEKTLIQENVLPQIHTVTGKTFMKNLFLASTRQNVGSYD 1260
QY 1261 GAYAPVLDPRSLNSTNTEKKTAAHFSKGEENLEGLGNQTKQIVEKYACTTRISPT 1320
Db 1261 GAYAPVLDPRSLNSTNTEKKTAAHFSKGEENLEGLGNQTKQIVEKYACTTRISPT 1320
QY 1321 SOONFVTOASKEALKOFLPLETELEKRIIVDDTSTQWSKNMCHLTPSTLTQIDVNEKE 1380
Db 1321 SOONFVTOASKEALKOFLPLETELEKRIIVDDTSTQWSKNMCHLTPSTLTQIDVNEKE 1380
QY 1381 KGAIQTSPSLDCITRSHSTPQANRSLPIAKVSSPSPRIPIYLTRVLFQDNSSHLPAAASY 1440
Db 1381 KGAIQTSPSLDCITRSHSTPQANRSLPIAKVSSPSPRIPIYLTRVLFQDNSSHLPAAASY 1440
QY 1441 RKKDSGVQSSHFLOQAKKNLSLAILTLENTGDQREVGLSATSNTSVYTKKVENTVLP 1500
Db 1441 RKKDSGVQSSHFLOQAKKNLSLAILTLENTGDQREVGLSATSNTSVYTKKVENTVLP 1500
QY 1501 KPDLFRTSKVELLPKWHYQKDLFPETESNGSPGHLDLVEGSLQGTGEGAIKWEANRP 1560
Db 1501 KPDLFRTSKVELLPKWHYQKDLFPETESNGSPGHLDLVEGSLQGTGEGAIKWEANRP 1560

QY 1561 CKVPPFLRVATESAKTPSKLLDLANDNHYGTQIPKEEWSQBSPEKTAFAKKDITLSL 1620
Db 1561 CKVPPFLRVATESAKTPSKLLDLANDNHYGTQIPKEEWSQBSPEKTAFAKKDITLSL 1620
QY 1621 NACESNHAAANEGQNKPEIEVTWAKQGRTERLCSQNPVLKRGHOREITRTTLOSQDOE 1680
Db 1621 NACESNHAAANEGQNKPEIEVTWAKQGRTERLCSQNPVLKRGHOREITRTTLOSQDOE 1680
QY 1681 IDYDDTISVEMKKEDEDIYDEDENQSPRSFQKTRHYFIAAVERLMDYGMSSSPHVLNR 1740
Db 1681 IDYDDTISVEMKKEDEDIYDEDENQSPRSFQKTRHYFIAAVERLMDYGMSSSPHVLNR 1740
QY 1741 AOSGSVPOEKVYVFOBTGOSTPOPLRGELNEHLGLGPYRAEVEDINMTFRNOASR 1800
Db 1741 AOSGSVPOEKVYVFOBTGOSTPOPLRGELNEHLGLGPYRAEVEDINMTFRNOASR 1800
QY 1801 PYSFYSSLSIYBEDQOQAEPRKQFVKPNETKTYPKVQVHMAPTKDEDFCKAWAYFSDV 1860
Db 1801 PYSFYSSLSIYBEDQOQAEPRKQFVKPNETKTYPKVQVHMAPTKDEDFCKAWAYFSDV 1860
QY 1861 DLEKOVHSGLIGPLAVCHTNTLNPAHGRQVTVQEPALFTIETDKSWYFTENMERNCA 1920
Db 1861 DLEKOVHSGLIGPLAVCHTNTLNPAHGRQVTVQEPALFTIETDKSWYFTENMERNCA 1920
QY 1921 PCNIQMEDPTKENYRFFHAINGYIMDTPLGLVMAQDQRIRWYLLSMGSHENIHSIHFSGH 1980
Db 1921 PCNIQMEDPTKENYRFFHAINGYIMDTPLGLVMAQDQRIRWYLLSMGSHENIHSIHFSGH 1980
QY 1981 VFTVRKKEBYKMAALYNLYPGVFETVEMLPKAGIWRVECLIGELHAGMSTFLVYVSNKC 2040
Db 1981 VFTVRKKEBYKMAALYNLYPGVFETVEMLPKAGIWRVECLIGELHAGMSTFLVYVSNKC 2040
QY 2041 QTPFLGMAASHIEDPQITASGGYQGNAPKLARLHYSGSINAWSTKEPFSWKVDLLAPMII 2100
Db 2041 QTPFLGMAASHIEDPQITASGGYQGNAPKLARLHYSGSINAWSTKEPFSWKVDLLAPMII 2100
QY 2101 HGKTKQARQKPSLYISQFIIMVSLDGKKWQTYGNSGTGLMVPFGNYDSSGKHNIEN 2160
Db 2101 HGKTKQARQKPSLYISQFIIMVSLDGKKWQTYGNSGTGLMVPFGNYDSSGKHNIEN 2160
QY 2161 PPIIARYIRLHPHTHYSIRSTLEMEIMGCDLNSCSMPLGMESKAI SDAQITASSYFTNMEA 2220
Db 2161 PPIIARYIRLHPHTHYSIRSTLEMEIMGCDLNSCSMPLGMESKAI SDAQITASSYFTNMEA 2220
QY 2221 TWSPKARLHLOGRSNARPOVNNPKWLQVDFQTKMKTGVTTCQVKSLLTSMVYKEPL 2280
Db 2221 TWSPKARLHLOGRSNARPOVNNPKWLQVDFQTKMKTGVTTCQVKSLLTSMVYKEPL 2280
QY 2281 ISSSODGHQWTLFFQNGKVKVFCGNQDSFTPVVNSLDPPLLTRYLRIHQSWVHQAIALRM 2340
Db 2281 ISSSODGHQWTLFFQNGKVKVFCGNQDSFTPVVNSLDPPLLTRYLRIHQSWVHQAIALRM 2340
QY 2341 EVLGCEAQDLY 2351
Db 2341 EVLGCEAQDLY 2351

RESULT 5

AAV52537

ID AAV52537 standard; protein; 2351 AA.

XX AAV52537;

XX AAV52537;

DT 28-FEB-2000 (first entry)

DE Human full-length factor VIII.

KW Factor VIII; haemophilia; proteolysis; heavy chain; light chain;

KW secretion; von Willebrand Factor; vWF; C2 domain; intron 22; inversion;

KW non-functional; bleeding disorder; coagulation; treatment.

OS Homo sapiens.

XX Key Location/Qualifiers
FH Peptide 1..19
FT /note= "signal peptide"
FT Protein 20..1708
FT /note= "Maximum length human factor VIII heavy chain"
FT Protein 1709..2351
FT /note= "Human factor VIII light chain"
XX WO959622-A1.
XX 25-NOV-1999.
XX 17-MAY-1999; 99WO-US010872.
XX 20-MAY-1999; 99US-00092000.
XX (GETH) GENENTECH INC.
FA Vemar GA;
XX W31; 2000-053195/04.
DR N-PSDB; AA238604.
XX New method for treating hemophilia A.
XX Disclosure; Fig 9; 50pp; English.
XX This sequence represents the full-length human factor VIII. In this form, the protein is 300 kD in size with the domain structure A1-A2-B-A3-C1-C2. However, prior to secretion, this 300 kD protein is proteolytically cleaved into a heavy chain (A1-A2-B, with continued proteolysis within the B domain resulting in molecules of varying length) and a light chain (A3-C1-C2) that remains non-covalently attached to the heavy chain. Upon secretion, factor VIII is rapidly cleared from the circulation unless it is bound by the plasma protein von Willebrand Factor (vWF). Factor VIII binds to vWF through the light chain, with known binding sites at the N-terminus and within the C2 domain. Haemophilia A is frequently caused by an intron 22 inversion in the factor VIII gene, which separates the gene into two sections, exons 1-22 becoming inverted and localised telomeric to the original site, while exons 23-26 remain isolated at the original site. Exons 23-26 a portion of the C1 and all of the C2 domains, without which factor VIII is non-functional. The factor VIII gene product of individuals with this mutation thus comprises domains A1-A2-B-A3 plus a fragment of C1, which on proteolysis is non-functional, resulting in a bleeding disorder. The invention relates to a novel method for for treating haemophilia A in a mammal able to produce the factor VIII heavy chain. The method comprises administering to the mammal either an effective amount of factor VIII light chain, or a gene encoding it, and may be useful for treating patients such as those whose haemophilia A is caused by intron 22 inversion. The recombinant factor VIII products of this invention are derived from well-characterised starting materials which therefore reduces the risk of pathogenic infection which was previously a problem when using donated plasma. Furthermore, the invention provides a more economic and potentially more effective treatment for haemophilia. There is also a need of providing factor VIII activity to patients who produce or are at risk of producing antibodies against full-length factor VIII
XX Sequence 2351 AA;
Query Match 99.9%; Score 12399; DB 3; Length 2351;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 2348; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
QY 1 MQELSTCFPLCLLRCSATRYIYLGAVELSDWYQSDLGELPVDARPPVPKSPFN 60
DB 1 MQELSTCFPLCLLRCSATRYIYLGAVELSDWYQSDLGELPVDARPPVPKSPFN 60
QY 61 TSVYVKTLFVEFTDHLFNIAKPRPPWMLGGLGTIOAEVYDVTWITLKNWASHFVSLHAV 120
DB 61 TSVYVKTLFVEFTDHLFNIAKPRPPWMLGGLGTIOAEVYDVTWITLKNWASHFVSLHAV 120

QY 121 GVSYWKASBAGYDDQTSOREKEDDKVPPGSSHTYVQVLKENGWASDPLCLTYLSLH 180
DB 121 GVSYWKASBAGYDDQTSOREKEDDKVPPGSSHTYVQVLKENGWASDPLCLTYLSLH 180
QY 181 VDLVKDLSGLIGALLVCREGSLAKEKTQTLHKFILLFAVDFEGSKWSHSETKNSLMQORD 240
DB 181 VDLVKDLSGLIGALLVCREGSLAKEKTQTLHKFILLFAVDFEGSKWSHSETKNSLMQORD 240
QY 241 AASARAWPKOHTWNGYVNRSLPGLIGCHRKSVYWHVIGMTTPEVHSIPLEGHTFLVRNH 300
DB 241 AASARAWPKOHTWNGYVNRSLPGLIGCHRKSVYWHVIGMTTPEVHSIPLEGHTFLVRNH 300
QY 301 ROASLEISPIETELTAQTLLMDLQFLLSCHISSHODGMEAVKVVDSPERPEPKWKNNE 360
DB 301 ROASLEISPIETELTAQTLLMDLQFLLSCHISSHODGMEAVKVVDSPERPEPKWKNNE 360
QY 361 EADYDDDLTDSMDVWRFDDNPSFIIQIRSVAKKHPTWVHYIAABEEDWDYAPLVLA 420
DB 361 EADYDDDLTDSMDVWRFDDNPSFIIQIRSVAKKHPTWVHYIAABEEDWDYAPLVLA 420
QY 421 PDRSYKSOYLNNGPQRIGRKYKVRPMAYTDETFKTRAIQHSGLIGLPLLYGEVDTL 480
DB 421 PDRSYKSOYLNNGPQRIGRKYKVRPMAYTDETFKTRAIQHSGLIGLPLLYGEVDTL 480
QY 481 LIIFXNQASRPYNIYPHGITDVRPLYSRRLPKGVKHLKDFPILPGEIPKYKWTVTVEDGP 540
DB 481 LIIFXNQASRPYNIYPHGITDVRPLYSRRLPKGVKHLKDFPILPGEIPKYKWTVTVEDGP 540
QY 541 TKSDPRCLTRYVSSPVNMRDIAAGLIGLPLLYCYKESVDKGNQINMSDKRNVLFSVDE 600
DB 541 TKSDPRCLTRYVSSPVNMRDIAAGLIGLPLLYCYKESVDKGNQINMSDKRNVLFSVDE 600
QY 601 NRSWLTENIOFLNPAGVQLEDPEFQASNIHMSINGYVFDLSQLSVCLHEVAWYILS 660
DB 601 NRSWLTENIOFLNPAGVQLEDPEFQASNIHMSINGYVFDLSQLSVCLHEVAWYILS 660
QY 661 IGAQTDFLSVFPGYTFKHKWVYEDTLTFPPSGTVPWMSMENGLWILGCHNSDFNRG 720
DB 661 IGAQTDFLSVFPGYTFKHKWVYEDTLTFPPSGTVPWMSMENGLWILGCHNSDFNRG 720
QY 721 MTALLKVSCKNKGVDYEDSIEDISAVLLSKNAIEPRFSQNSRHPSTROKQFNATTI 780
DB 721 MTALLKVSCKNKGVDYEDSIEDISAVLLSKNAIEPRFSQNSRHPSTROKQFNATTI 780
QY 781 PENDIEKTDPEFAHTEPMPKIONVSSDLMLLRQSPHGLSLDLOBAKYETESDDPS 840
DB 781 PENDIEKTDPEFAHTEPMPKIONVSSDLMLLRQSPHGLSLDLOBAKYETESDDPS 840
QY 841 PGASIDNNSLSMTHTFRPQLHHSQDGVFTPESSGLQLRNEKLGTTAATLKKLDFKVSST 900
DB 841 PGASIDNNSLSMTHTFRPQLHHSQDGVFTPESSGLQLRNEKLGTTAATLKKLDFKVSST 900
QY 901 SNNLSTIPSDMLAAGTNTSSLGPPMPVHYVDSQDITLFGKKSPLTESGGPISLSEE 960
DB 901 SNNLSTIPSDMLAAGTNTSSLGPPMPVHYVDSQDITLFGKKSPLTESGGPISLSEE 960
QY 961 NNDKLLBSGLMNSQESSWKNVSTESGRLPKGRAGHPALLTKDNALFKVSIILLKTN 1020
DB 961 NNDKLLBSGLMNSQESSWKNVSTESGRLPKGRAGHPALLTKDNALFKVSIILLKTN 1020
QY 1021 KTSNNSATNRKTHIDGPSLLIENSPPVQNILESDETFKKTPTLIDRLMDKNATLRL 1080
DB 1021 KTSNNSATNRKTHIDGPSLLIENSPPVQNILESDETFKKTPTLIDRLMDKNATLRL 1080
QY 1081 NEMSNKTTSSKNMNVQKKEGPIPPDAQNDPMSFFKMLFLPESARWQRTHQKSLNSG 1140
DB 1081 NEMSNKTTSSKNMNVQKKEGPIPPDAQNDPMSFFKMLFLPESARWQRTHQKSLNSG 1140
QY 1141 QGPSKQLVSLGPEKSVGQNFLESKNVVVGKGFETKDVGLKMWVFPSSRNLFITNLDN 1200
DB 1141 QGPSKQLVSLGPEKSVGQNFLESKNVVVGKGFETKDVGLKMWVFPSSRNLFITNLDN 1200
QY 1201 LHENNTNQEKIQEEIEKKTETLIQENVVLPQIHTVTGTQNFPMKNLFLITRQNVGSGYD 1260

Db 1201 LHENNTNQEKKIQSEIBKKEKTLIQENVVLPQIHVTGKTGKPMKXNLFLLSTRQNVESYD 1260
QY 1261 GAYAVLODFSLNDSNTRTKKHTAHFSPKGBEENLGLGNQTKQIVKVIACHTTISNT 1320
Db 1261 GAYAVLODFSLNDSNTRTKKHTAHFSPKGBEENLGLGNQTKQIVKVIACHTTISNT 1320
QY 1321 SQONFVQSRKALQFRLPDETELEKRIIVDDTSTQWSNMKHLTPSTLTQIDYNEKE 1380
Db 1321 SQONFVQSRKALQFRLPDETELEKRIIVDDTSTQWSNMKHLTPSTLTQIDYNEKE 1380
QY 1381 KGAITQSPSLDCLTRSHSIPQANRSPPLPIAKVSSPFSIRPIYLTAVLPQDNSSHLPAASY 1440
Db 1381 KGAITQSPSLDCLTRSHSIPQANRSPPLPIAKVSSPFSIRPIYLTAVLPQDNSSHLPAASY 1440
QY 1441 RKDSGVQESSHFLQAGKKNLSAILLLEMTGDQREVGSGLSATNSVTYKKEVNTVLP 1500
Db 1441 RKDSGVQESSHFLQAGKKNLSAILLLEMTGDQREVGSGLSATNSVTYKKEVNTVLP 1500
QY 1501 KPDLKPTSGVLLPKVHIYQKDLPTTETSNGSPGHLDLVEGSLQGTGEGAIKWNENRP 1560
Db 1501 KPDLKPTSGVLLPKVHIYQKDLPTTETSNGSPGHLDLVEGSLQGTGEGAIKWNENRP 1560
QY 1561 GKVPFLRVATESAKTPSKLLDPLAWDNHYGTQIPKEEWSQESKPEKTAFFKKDTILSL 1620
Db 1561 GKVPFLRVATESAKTPSKLLDPLAWDNHYGTQIPKEEWSQESKPEKTAFFKKDTILSL 1620
QY 1621 NACESNHAIAINEGKNPEIEVTWAKQRTSRLCSQNPVPLKRCREITRTTLOSQDOE 1680
Db 1621 NACESNHAIAINEGKNPEIEVTWAKQRTSRLCSQNPVPLKRCREITRTTLOSQDOE 1680
QY 1681 IDYDITISVEMKKEPDIYDENQSPRSFQKTRHYFIAAVERLWDYGMSSSPHVLNR 1740
Db 1681 IDYDITISVEMKKEPDIYDENQSPRSFQKTRHYFIAAVERLWDYGMSSSPHVLNR 1740
QY 1741 AQSGVPOPKVVFQFTDGSSTQPLRYGELNEHLGLGPIYRAEVEDINIMVTFRNQASR 1800
Db 1741 AQSGVPOPKVVFQFTDGSSTQPLRYGELNEHLGLGPIYRAEVEDINIMVTFRNQASR 1800
QY 1801 PYSFYSSILSYEDORQGAEPKPNFVKNETTYFWKVOHMAPTKDFDCKAWAYFSDV 1860
Db 1801 PYSFYSSILSYEDORQGAEPKPNFVKNETTYFWKVOHMAPTKDFDCKAWAYFSDV 1860
QY 1861 DLEKDVHSLGLPLVCHTNTLNPAHGRQVTVQEFALPTTIFDETKSWYFTENMERNCRA 1920
Db 1861 DLEKDVHSLGLPLVCHTNTLNPAHGRQVTVQEFALPTTIFDETKSWYFTENMERNCRA 1920
QY 1921 PCNIQWEDPTPKENYRFHAINGYIMDTPLGLYMAQDQRIKRYLLSGNSNENHSIHFSGH 1980
Db 1921 PCNIQWEDPTPKENYRFHAINGYIMDTPLGLYMAQDQRIKRYLLSGNSNENHSIHFSGH 1980
QY 1981 VFTVRKKEBKVALYNLYPGVFETVEMLPKAGIWRVECLIGEHLAGMSTFLVYSNKC 2040
Db 1981 VFTVRKKEBKVALYNLYPGVFETVEMLPKAGIWRVECLIGEHLAGMSTFLVYSNKC 2040
QY 2041 QTPLGVASGHIRDFQITAGQYQWAPKLARLHYSGSINAWSTKPFWSIKVDLIAPMII 2100
Db 2041 QTPLGVASGHIRDFQITAGQYQWAPKLARLHYSGSINAWSTKPFWSIKVDLIAPMII 2100
QY 2101 HGIKTQGRQKPSLSYISOFIIMYSLDGKKWOTYRGNSTGTLWVFGNVYDSSGIKHNEN 2160
Db 2101 HGIKTQGRQKPSLSYISOFIIMYSLDGKKWOTYRGNSTGTLWVFGNVYDSSGIKHNEN 2160
QY 2161 PPIIARYIRLHPHTHSIRSTRLMELMGCDLNSCMPLGMESKAISDAQITASSYFTNMFA 2220
Db 2161 PPIIARYIRLHPHTHSIRSTRLMELMGCDLNSCMPLGMESKAISDAQITASSYFTNMFA 2220
QY 2221 TWSPPSKARLHLQGRSNAPQVANNPKWLQVDPOKTKMYGTVTQGVKSLLTSMYKKEFL 2280
Db 2221 TWSPPSKARLHLQGRSNAPQVANNPKWLQVDPOKTKMYGTVTQGVKSLLTSMYKKEFL 2280
QY 2281 ISSODGHOHTLFFQNGKVKVQGNQDSFTPVVNSLDPLLTLYRIHQPQSWHQIALRM 2340

Db 2281 ISSODGHOHTLFFQNGKVKVQGNQDSFTPVVNSLDPLLTLYRIHQPQSWHQIALRM 2340
QY 2341 EVLGCEAODLY 2351
Db 2341 EVLGCEAODLY 2351
RESULT 6
ABB99528
ID ABB99528 standard; protein; 2351 AA.
XX
AC ABB99528;
XX
DT 03-MAR-2003 (first entry)
XX
DE Amino acid sequence of human Factor VIII.
XX
KW Factor VIII; non-viral vesicle vector; vesicular membrane; hepatitis B; envelope protein; Factor IX; liver cell; gene therapy; haemophilia.
XX
OS Homo sapiens.
XX
PN WO200286091-A2.
XX
PD 31-OCT-2002.
XX
PF 25-APR-2002; 2002WO-US013164.
XX
PR 25-APR-2001; 2001US-0286314P.
XX
PA (REGC) UNIV CALIFORNIA.
XX
PI Chien KR, Hoshijima M;
DR WPI; 2003-093125/08.
DR N-PSDB; ABV77052.
XX
PT New nor-viral vesicle vector comprises vesicular membrane with hepatitis B envelope protein and nucleic acid expression construct comprising complete factor VIII or IX coding sequence, useful for treating hemophilia.
XX
PS Disclosure; Page 22-32; 34pp; English.
XX
The present sequence represents a human Factor VIII. Factor VIII polynucleotides are used to construct non-viral vesicle vectors. These vectors comprise a vesicular membrane with hepatitis B envelope protein exposed on the vesicle surface, and a nucleic acid expression construct comprising a complete factor VIII or factor IX coding sequence and a promoter sequence functional in liver cells. The non-viral vesicle vector is useful in gene therapy protocols for treating haemophilia
XX
SQ Sequence 2351 AA;
Query Match 99.9%; Score 12399; DB 6; Length 2351;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 2348; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
QY 1 MOIELSTCFFCLLRFCFSATRRYYLGAVELSWDVMQSDLGELPVDARPPRPVKSFPFN 60
Db 1 MOIELSTCFFCLLRFCFSATRRYYLGAVELSWDVMQSDLGELPVDARPPRPVKSFPFN 60
QY 61 TSVVYKTLFVEFTDHLFNIAKRPFPNMGLLGPITIAEVYDVTITLKNMASHPVSLHAV 120
Db 61 TSVVYKTLFVEFTDHLFNIAKRPFPNMGLLGPITIAEVYDVTITLKNMASHPVSLHAV 120
QY 121 GVSYYKASGEGAEYDDQTSQRKEDDKVFPGSGHTYVWVLKENGPMASDPLCLTYSYLSH 180
Db 121 GVSYYKASGEGAEYDDQTSQRKEDDKVFPGSGHTYVWVLKENGPMASDPLCLTYSYLSH 180
QY 181 VDLVKDLNSGLIGALLVCRGSLAKEKTQTLHKFTLLFAVDEGKSWHSETKNSLMQDRD 240
Db 181 VDLVKDLNSGLIGALLVCRGSLAKEKTQTLHKFTLLFAVDEGKSWHSETKNSLMQDRD 240

RESULT 7

AAW11416
ID AAW11416 standard; protein; 2351 AA.
XX AAW11416;
AC AAW11416;
XX 20-NOV-1997 (first entry)
XX Active Factor VIII:C analogue A774X.
XX Factor VIII:C; analogue; glycoprotein; blood coagulation cascade;
KW fibrinogen; fibrin clot; haemostasis; haemophilia A; bleeding diathesis;
KW plasma protease; thrombin; immunogen; antibody; haemophilic; therapy;
KW proteolytic cleavage.
XX Homo sapiens.
OS Synthetic.
XX Key Location/Qualifiers
FH . . 19
FT Peptide /note= "signal peptide"
FT Protein 20..2351
FT Region /note= "mature Factor VIII:C"
FT /note= "heavy chain fragment"
FT Domain 760..1667
FT /note= "B domain"
FT Modified-site 774
FT /label= phe, Glu, Pro
FT Region 1668..2350
FT /note= "light chain fragment"
XX WO9703195-A1.
XX 30-JAN-1997.
XX 09-JUL-1996; 96WO-US011444.
XX 11-JUL-1995; 95US-0001025P.
XX (CHIR) CHIRON CORP.
XX Hung DT, Cohen FE, Innis M;
XX WPI; 1997-119050/11.
XX Factor VIII:C analog modified adjacent to a non-activating Arg residue -
XX used in the treatment of haemophilias, by improvement of haemostasis.
XX Claim 26; Page: 90pp; English.
XX AAW11330-W11472 represent active Factor VIII:C analogues of the
XX invention. These sequences were created by mutating the wild type factor
XX VIII:C coding sequence (see AAW11357) using mutagenic primers. The
XX analogues comprise a native Factor VIII:C polypeptide modified at a site
XX adjacent to a non-activating Arg residue so that a Arg-Pro or Pro-Arg
XX dipeptide is created. Factor VIII:C is a large glycoprotein that
XX participates in the blood coagulation cascade that ultimately converts
XX soluble fibrinogen to insoluble fibrin clot, effecting haemostasis. A
XX deficiency in Factor VIII:C is responsible for haemophilia A, which is an
XX X-chromosome-linked inherited bleeding diathesis. Factor VIII:C is
XX activated by plasma proteases, such as thrombin. During activation the
XX mature polypeptide is cleaved to generate heavy and light chain fragments
XX that are further cleaved. Complexes of two or more of the analogues,
XX nucleic acids and vectors encoding them may be used alone or in
XX conjunction with each other, for the prevention or treatment of active
XX Factor VIII:C deficiency in a mammal. The analogues may be used as
XX immunogens to raise antibodies, and in the treatment of haemophilias, by
XX improvement of haemostasis. The analogues are resistant to proteolytic
XX cleavage and display increased plasma half-life. They may be administered
XX at lower dosages and by different modes of administration
XX Sequence 2351 AA;

Query Match	99.8%;	Score 12395;	DB 2;	Length 2351;
Best Local Similarity	99.8%;	Pred. No. 0;		
Matches 2347;	Conservative 1;	Mismatches 3;	Indels 0;	Gaps 0;
QY	1	MQIELSTCFELCLRFCSATRRYYIGAVELSDWYMQSDLGELPVDARPPRPVKSPFFN	60	
DB	1	MQIELSTCFELCLLRFCFSATRRYYIGAVELSDWYMQSDLGELPVDARPPRPVKSPFFN	60	
QY	51	TSVYVKTLFVEFTDHLFNIAKPRPPWMLGLOPTTQAEVYDTVITLKNMASHPVSLEAV	120	
DB	51	TSVYVKTLFVEFTDHLFNIAKPRPPWMLGLOPTTQAEVYDTVITLKNMASHPVSLEAV	120	
QY	121	GVSYWKASEGAEYDDOTSQREKEDDKVPFGSHTVWQVLKENGPMASDPLCLTYSYLH	180	
DB	121	GVSYWKASEGAEYDDOTSQREKEDDKVPFGSHTVWQVLKENGPMASDPLCLTYSYLH	180	
QY	181	VDLVKDLNSGLIGALLVCREGLAKEKTQTLHKFILLFAVDEGKSWHSETKNSLMQDRD	240	
DB	181	VDLVKDLNSGLIGALLVCREGLAKEKTQTLHKFILLFAVDEGKSWHSETKNSLMQDRD	240	
QY	241	AASARAWPMGHTVNGYVNRSLPGLIGCHRKSVYWHVIGMTTPEVHSIFLEHTFLVRNH	300	
DB	241	AASARAWPMGHTVNGYVNRSLPGLIGCHRKSVYWHVIGMTTPEVHSIFLEHTFLVRNH	300	
QY	301	ROASLEISPIITFLTAQTLLMDLGOFLLSCHISSHOHDGMEAYVKYVDSCEPBPOLMKNE	360	
DB	301	ROASLEISPIITFLTAQTLLMDLGOFLLRCHISSHOHDGMEAYVKYVDSCEPBPOLMKNE	360	
QY	361	EAEDYDDDLTDSMDVVRPDDNSPFTQIRSAVKGPKTWVHYIAABEEDWDYAPLVLA	420	
DB	361	EAEDYDDDLTDSMDVVRPDDNSPFTQIRSAVKGPKTWVHYIAABEEDWDYAPLVLA	420	
QY	421	PDORSYKSOYLNNGPORIGRKVKYKRMAYDETEKTEATQHESGILGLLYGVGDTL	480	
DB	421	PDORSYKSOYLNNGPORIGRKVKYKRMAYDETEKTEATQHESGILGLLYGVGDTL	480	
QY	481	LIIFKNQASRPYNIYPHGITDVRPLYSRRLPKGVKHLKDFPILGCEIFKYKWTVTVEGDP	540	
DB	481	LIIFKNQASRPYNIYPHGITDVRPLYSRRLPKGVKHLKDFPILGCEIFKYKWTVTVEGDP	540	
QY	541	TKSDPRCLTRYYSFVNMRDLASGLIGPLLCYKESVDQKGNQIMSKRNVILFSVFDE	600	
DB	541	TKSDPRCLTRYYSFVNMRDLASGLIGPLLCYKESVDQKGNQIMSKRNVILFSVFDE	600	
QY	601	NESWLTNTIORFLNPAGVQLEDPEPCASNTMHSINGYVDSLSQLSVCLHEVAYWYLS	660	
DB	601	NRSWLTNTIORFLNPAGVQLEDPEPCASNTMHSINGYVDSLSQLSVCLHEVAYWYLS	660	
QY	661	IGAQTDFLSVFPFGYTFKHKVYEDTLTFPPSGTVPMSMENPGLWILGCHNSDFNRNG	720	
DB	661	IGAQTDFLSVFPFGYTFKHKVYEDTLTFPPSGTVPMSMENPGLWILGCHNSDFNRNG	720	
QY	721	MTALLKVSCKDNKTDYEDSDYEDISAYLLSKNNAIEPRSPSONSRHPSTKQKNATTI	780	
DB	721	MTALLKVSCKDNKTDYEDSDYEDISAYLLSKNNAIEPRSPSONSRHPSTKQKNATTI	780	
QY	781	PENDTEKTDPMFAETPMKTCNVSSSLLMLROSPTPHGLSLDLOEAKVETSDOPS	840	
DB	781	PENDLEKTDPMFAETPMKTCNVSSSLLMLROSPTPHGLSLDLOEAKVETSDOPS	840	
QY	841	PGAIDSNNSLSEMTFRPQLHSGDMVTPPSGQLRLNEXLGTATAELKKLDPKVST	900	
DB	841	PGAIDSNNSLSEMTFRPQLHSGDMVTPPSGQLRLNEXLGTATAELKKLDPKVST	900	
QY	901	SNNLISITPSDNLAAGTNTSSLPSPMPVYDSDTLTGKXSSPLTESGGPLSGEE	960	
DB	901	SNNLISITPSDNLAAGTNTSSLPSPMPVYDSDTLTGKXSSPLTESGGPLSGEE	960	
QY	961	NNDKSLLESGLMNSQESSGNVNSTESGRLFKGRAGHPALLTKDNALPKVSTSLKTN	1020	
DB	961	NNDKSLLESGLMNSQESSGNVNSTESGRLFKGRAGHPALLTKDNALPKVSTSLKTN	1020	

QY 1021 KTSNRSATNRKTHIDGPSLLIENSFVWQNILESDETEPKKVTPLIHDRMLDKNATLRL 1080
DB 1021 KTSNRSATNRKTHIDGPSLLIENSFVWQNILESDETEPKKVTPLIHDRMLDKNATLRL 1080
QY 1081 NHMSKNTSSKNMVAQKKEGPIPPDAQNPDMSPFKMLFLPESARWIQRTGKNSLNG 1140
DB 1081 NHMSKNTSSKNMVAQKKEGPIPPDAQNPDMSPFKMLFLPESARWIQRTGKNSLNG 1140
QY 1141 QGSPKQLVSLGPEKSEVEGQNFISEKNKVVVGKBEFTKDVGKEMVFPSSRNLFITNLND 1200
DB 1141 QGSPKQLVSLGPEKSEVEGQNFISEKNKVVVGKBEFTKDVGKEMVFPSSRNLFITNLND 1200
QY 1201 LHENNTNHOEKKIQEBIEKKGTILQENVVLPOIHTVTG*KNFMKNLFLLLSTRQNVESGYD 1260
DB 1201 LHENNTNHOEKKIQEBIEKKGTILQENVVLPOIHTVTG*KNFMKNLFLLLSTRQNVESGYD 1260
QY 1261 GAYAPVLODFRSLNDSTNRTKKTHTAHFSKKGEEENLEGHGNOTKQIVEKYACTTISPN 1320
DB 1261 GAYAPVLODFRSLNDSTNRTKKTHTAHFSKKGEEENLEGHGNOTKQIVEKYACTTISPN 1320
QY 1321 SQCNFVTRSKALQKFRPLPLETELEKRIIVDDTSTQSKNMKHLTPSTLTQIDYNEKE 1380
DB 1321 SQCNFVTRSKALQKFRPLPLETELEKRIIVDDTSTQSKNMKHLTPSTLTQIDYNEKE 1380
QY 1381 KGAIQTSPSLDCILTRSHSIPQANRSPPIAKVSSPPSIRPIYLVTRVLPQDNSSHLPAA 1440
DB 1381 KGAIQTSPSLDCILTRSHSIPQANRSPPIAKVSSPPSIRPIYLVTRVLPQDNSSHLPAA 1440
QY 1441 RKQSGVQSSHFLOCAKKNLISLALTLEMTGDQREVSGISATNSVTYKVENVTLP 1500
DB 1441 RKQSGVQSSHFLOCAKKNLISLALTLEMTGDQREVSGISATNSVTYKVENVTLP 1500
QY 1501 KPDLPKTSKVELLPKVHIYQKDLFFETSNQSPGHLDVEGSLLOGTSCAIKWEANRP 1560
DB 1501 KPDLPKTSKVELLPKVHIYQKDLFFETSNQSPGHLDVEGSLLOGTSCAIKWEANRP 1560
QY 1561 GKVPFLRVATSSAKTPSKLLDPLANDNHYGQIPKEWKSGQKSPKTAFAKKOTILSL 1620
DB 1561 GKVPFLRVATSSAKTPSKLLDPLANDNHYGQIPKEWKSGQKSPKTAFAKKOTILSL 1620
QY 1621 NACESNHAIAANEGONKPEIEVTWAKQRTKELCSQNPVLKSHOREITRTTLQSDQEE 1680
DB 1621 NACESNHAIAANEGONKPEIEVTWAKQRTKELCSQNPVLKSHOREITRTTLQSDQEE 1680
QY 1681 IDYDDTISVEMKKEFDIYDEDENQSPRSFQKTRHYFIAAVERLWDYGMSSSPHYLRNR 1740
DB 1681 IDYDDTISVEMKKEFDIYDEDENQSPRSFQKTRHYFIAAVERLWDYGMSSSPHYLRNR 1740
QY 1741 AQSGSVQPKVQKVVQFTDGSFTQPLRYGELNEHGLGPIYRAEYEDNIMVTFRQASR 1800
DB 1741 AQSGSVQPKVQKVVQFTDGSFTQPLRYGELNEHGLGPIYRAEYEDNIMVTFRQASR 1800
QY 1801 PYSFYSLIASYEDDQQAEPKRNFKVKNETKTYFWKVQHNHNAPTXDEFDCRWAYFSDV 1860
DB 1801 PYSFYSLIASYEDDQQAEPKRNFKVKNETKTYFWKVQHNHNAPTXDEFDCRWAYFSDV 1860
QY 1861 DLEKDVHSGLIGLPLVCHNTNLNPAHGRQVTVQEFALFTTIFDETKSWYFTENMERNCRA 1920
DB 1861 DLEKDVHSGLIGLPLVCHNTNLNPAHGRQVTVQEFALFTTIFDETKSWYFTENMERNCRA 1920
QY 1921 PCNIQMEDTFEKNYRFAHNGVIMDTLPLGVNAQDQRIWYLLSNGSNENIHSIFPSGH 1980
DB 1921 PCNIQMEDTFEKNYRFAHNGVIMDTLPLGVNAQDQRIWYLLSNGSNENIHSIFPSGH 1980
QY 1981 VFTVRKKEEYKALYNLYPGVFTVEMLPKAGINRVECLIGSHLAGMSTLFLVYSNKC 2040
DB 1981 VFTVRKKEEYKALYNLYPGVFTVEMLPKAGINRVECLIGSHLAGMSTLFLVYSNKC 2040
QY 2041 QTPPLGNASGHIRDFQITASQYQGWAPKARLHYSIGSINAWSTKEPFSWIKVDLLAPM 2100
DB 2041 QTPPLGNASGHIRDFQITASQYQGWAPKARLHYSIGSINAWSTKEPFSWIKVDLLAPM 2100
QY 2101 HGKTKQARQKFSLSLTSQFIIMYSLDGKKQTYRGNSIGTLMVFFGNVDSSGKINIFN 2160

DB 2101 HGKTKQARQKFSLSLTSQFIIMYSLDGKKQTYRGNSIGTLMVFFGNVDSSGKINIFN 2160
QY 2161 PPIIARYIRLPHTHYSIRSTIRLMELMGCDLNSCMPLGMSKASDAQITASSYFTNMPA 2220
DB 2161 PPIIARYIRLPHTHYSIRSTIRLMELMGCDLNSCMPLGMSKASDAQITASSYFTNMPA 2220
QY 2221 TWSPSKALHLHQRSNAWRPQVNNPKFWLQVDFQKTMKVTGTTQGVKSLTSMYVKEFL 2280
DB 2221 TWSPSKALHLHQRSNAWRPQVNNPKFWLQVDFQKTMKVTGTTQGVKSLTSMYVKEFL 2280
QY 2281 ISSSODGQHWTLFFQNGKVKVQFNQDSFTFVNSLDPLLTTRYLRHPQSWHQAIALRM 2340
DB 2281 ISSSODGQHWTLFFQNGKVKVQFNQDSFTFVNSLDPLLTTRYLRHPQSWHQAIALRM 2340
QY 2341 EVLGCERQDLY 2351
DB 2341 EVLGCERQDLY 2351
RESULT 8
AAW11343
ID AAW11343 standard; protein; 2351 AA.
XX AC AAW11343;
XX DT 17-NOV-1997 (first entry)
XX DE Active Factor VIII:C analogue S224X.
XX KW Factor VIII:C; analogue; glycoprotein; blood coagulation cascade;
XX KW fibrinogen; fibrin clot; haemostasis; haemophilia A; bleeding diathesis;
XX KW plasma protease; thrombin; immunogen; antibody; haemophiliac; therapy;
XX XX proteolytic cleavage.
OS Homo sapiens.
OS Synthetic.
XX FH Key Location/Qualifiers
FT Peptide 1..19 /note= "signal peptide"
FT Protein 20..2351 /note= "mature Factor VIII:C"
FT Region 20..1667 /note= "heavy chain fragment"
FT Modified-site 243 /label= Phe, Glu, Pro
FT Domain 760..1667 /note= "B domain"
FT Region 1668..2350 /note= "light chain fragment"
XX MO9703195-A1.
XX PN 30-JAN-1997.
XX PD 09-JUL-1996; 96WO-US011444.
XX PF 11-JUL-1995; 95US-00010259.
XX PR (CHIR) CHIRON CORP.
XX PA Hung DT, Cohen FE, Innis M;
XX PI WPI; 1997-119050/11.
XX DR Factor VIII:C analog modified adjacent to a non-activating Arg residue -
XX PT used in the treatment of haemophiliacs, by improvement of haemostasis.
XX PS Claim 10; Page; 90pp; English.
XX CC AAW11330-W11472 represent active Factor VIII:C analogues of the
invention. These sequences were created by mutating the wild type Factor

VIII:C coding sequence (see AAT51357) using mutagenic primers. The analogues comprise a native Factor VIII:C polypeptide modified at a site adjacent to a non-activating Arg residue so that a Arg-Pro or Pro-Arg dipeptide is created. Factor VIII:C is a large glycoprotein that participates in the blood coagulation cascade that ultimately converts soluble fibrinogen to insoluble fibrin clot, effecting haemostasis. A deficiency in Factor VIII:C is responsible for haemophilia A, which is an X-chromosome-linked inherited bleeding diathesis. Factor VIII:C is activated by plasma proteases, such as thrombin. During activation the mature polypeptide is cleaved to generate heavy and light chain fragments that are further cleaved. Complexes of two or more of the analogues, nucleic acids and vectors encoding them may be used alone or in conjunction with each other, for the prevention or treatment of active Factor VIII:C deficiency in a mammal. The analogues may be used as immunogens to raise antibodies, and in the treatment of haemophilias, by improvement of haemostasis. The analogues are resistant to proteolytic cleavage and display increased plasma half-life. They may be administered at lower dosages and by different modes of administration

XX Sequence 2351 AA;

Query March 99.8%; Score 12395; DB 2; Length 2351;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 2347; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 MQEILSTCFELCLLRCFCFSATRRYYLGAVELSDYMQSDLGELPVDARPPRPVPSFPPN 60
DB 1 MQEILSTCFELCLLRCFCFSATRRYYLGAVELSDYMQSDLGELPVDARPPRPVPSFPPN 60

QY 61 TSVVVKTLFVETFDHLENIAPRPFPWMLGLETQAEVYDVTVITLKNMASHPVSLHAV 120
DB 61 TSVVVKTLFVETFDHLENIAPRPFPWMLGLETQAEVYDVTVITLKNMASHPVSLHAV 120

QY 121 GVSYNKASGAYDQTSOREKDDKVRPGGSHYVQVLKENGPMASPLCLTYSLSH 180
DB 121 GVSYNKASGAYDQTSOREKDDKVRPGGSHYVQVLKENGPMASPLCLTYSLSH 180

QY 181 VDLVKDLSGLIGALLVCREGSLAKEKTQTLHKFILLFAVPDEGKSWHSETKNSLMQDRD 240
DB 181 VDLVKDLSGLIGALLVCREGSLAKEKTQTLHKFILLFAVPDEGKSWHSETKNSLMQDRD 240

QY 241 AASARAWPKMTVNGVKNRSLPGLIGCHRKSVYVHVGMTTPVHSIFLEGHTFLVRNH 300
DB 241 AASARAWPKMTVNGVKNRSLPGLIGCHRKSVYVHVGMTTPVHSIFLEGHTFLVRNH 300

QY 301 ROASLEISPIITLTACTLMDLQGLLSCHISSHQHDGMEAYVKVDSCPEEPQLRMKNE 360
DB 301 ROASLEISPIITLTACTLMDLQGLLSCHISSHQHDGMEAYVKVDSCPEEPQLRMKNE 360

QY 361 EAEDYDDDLTDSEMDVVRPDDDNSPSFIQIRSVAKKHPTWVHYTAAEBEDWDYAPLVLA 420
DB 361 EAEDYDDDLTDSEMDVVRPDDDNSPSFIQIRSVAKKHPTWVHYTAAEBEDWDYAPLVLA 420

QY 421 PDRSVKSYLNNQPGRIQKVKYRPMAYTDETFKTREAIQHSIGILPLYGEVGDTL 480
DB 421 PDRSVKSYLNNQPGRIQKVKYRPMAYTDETFKTREAIQHSIGILPLYGEVGDTL 480

QY 481 LIIFKNOASRPNIYPHGIITVRLPYSRRLPKGVKHLKDFPILPGBIPKYKMTVTVEDGP 540
DB 481 LIIFKNOASRPNIYPHGIITVRLPYSRRLPKGVKHLKDFPILPGBIPKYKMTVTVEDGP 540

QY 541 TKSDPRCLTRYSSFFVNMRDLASGLIGELLICYKESVDQGNQIMSKRNVLIFSVPDE 600
DB 541 TKSDPRCLTRYSSFFVNMRDLASGLIGELLICYKESVDQGNQIMSKRNVLIFSVPDE 600

QY 601 NRSWYLTENIQRLNPAGVQLEDPEFQASNMHNSINGVVFSLQLSVCLHEVAVYILS 660
DB 601 NRSWYLTENIQRLNPAGVQLEDPEFQASNMHNSINGVVFSLQLSVCLHEVAVYILS 660

QY 661 IGAQTDFLSVFSGYTFKHKMYEDTLTLPFSGSETVMSMENPGLWILGCHNSDFNRG 720
DB 661 IGAQTDFLSVFSGYTFKHKMYEDTLTLPFSGSETVMSMENPGLWILGCHNSDFNRG 720

QY 721 MTALLKVSQCDKNTGYYEDSDYEDISAYLLSKNNAIEPRSPQNSRHPSTROKQFNATTI 780
DB 721 MTALLKVSQCDKNTGYYEDSDYEDISAYLLSKNNAIEPRSPQNSRHPSTROKQFNATTI 780

QY 781 PNDIEKTPPFAHRTMPKIQNVSSDILLMLLRQSPTPHGLSLDLQEAKEYTSDDFS 840
DB 781 PNDIEKTPPFAHRTMPKIQNVSSDILLMLLRQSPTPHGLSLDLQEAKEYTSDDFS 840

QY 841 PCAIDSNLSLSEMTHERPQLHSGDMVTPESGLQRLNEKLGTTAATLKKLDFKVSST 900
DB 841 PCAIDSNLSLSEMTHERPQLHSGDMVTPESGLQRLNEKLGTTAATLKKLDFKVSST 900

QY 901 SNNLSTTPSDNLAAGTNTSSLGPPSPMPVHYDSQDITLFGKKSPSPITESGGPISLSE 960
DB 901 SNNLSTTPSDNLAAGTNTSSLGPPSPMPVHYDSQDITLFGKKSPSPITESGGPISLSE 960

QY 961 NNDKSLLESGLMNSQESSGKRVNVESTESGRLPFGKRAHGPALLTKDNALFKVISILKTN 1020
DB 961 NNDKSLLESGLMNSQESSGKRVNVESTESGRLPFGKRAHGPALLTKDNALFKVISILKTN 1020

QY 1021 KTSNNSATNRKTHIDGPSLLIENSPSVMQNTILSDTEPKVTPLTHDMMDKNATALRL 1080
DB 1021 KTSNNSATNRKTHIDGPSLLIENSPSVMQNTILSDTEPKVTPLTHDMMDKNATALRL 1080

QY 1081 NHMSNKTSSKNMEMVQKKGGPIPPDAQNPDMSFFKMLFLPESARWQRTHGKNSLNSG 1140
DB 1081 NHMSNKTSSKNMEMVQKKGGPIPPDAQNPDMSFFKMLFLPESARWQRTHGKNSLNSG 1140

QY 1141 QGPSKQLVSLGPEKSVGQNFLESKNKVVGKGGFTKDVGLKEMVFPSSRNLFITNLDN 1200
DB 1141 QGPSKQLVSLGPEKSVGQNFLESKNKVVGKGGFTKDVGLKEMVFPSSRNLFITNLDN 1200

QY 1201 LHENNTHQEKKIQEEIEKKETLQENVVLPOIHTVGTGKNFMKNLFLSLTRQNVGSD 1260
DB 1201 LHENNTHQEKKIQEEIEKKETLQENVVLPOIHTVGTGKNFMKNLFLSLTRQNVGSD 1260

QY 1261 GAYAPVLQDFRSLNDSTNRKKTAFHFKKGBEENLEGIGNOTKQIYEKYACTTISPT 1320
DB 1261 GAYAPVLQDFRSLNDSTNRKKTAFHFKKGBEENLEGIGNOTKQIYEKYACTTISPT 1320

QY 1321 SQQNVTVQSSKRALQFRLPIEETLEKRIIVDDTSTOWSKNMKHLTSTLTQIDYNEKE 1380
DB 1321 SQQNVTVQSSKRALQFRLPIEETLEKRIIVDDTSTOWSKNMKHLTSTLTQIDYNEKE 1380

QY 1381 KCAITQSPLSDCLTSHSIPOANRSPLEIAKVSSPESIRPIYLTVLFPQDNSSHUPAASY 1440
DB 1381 KCAITQSPLSDCLTSHSIPOANRSPLEIAKVSSPESIRPIYLTVLFPQDNSSHUPAASY 1440

QY 1441 RKKDSGVQESSHFLQGAKKNNLSLAILTEMTGDQREYGSIGTSATNSVTYKKVENTVLP 1500
DB 1441 RKKDSGVQESSHFLQGAKKNNLSLAILTEMTGDQREYGSIGTSATNSVTYKKVENTVLP 1500

QY 1501 XPDLPKTSKGVLELIPKVIHYKDLFPPTTSNGSPGHLDLVEGSLLOQTGEGAIKWEANRP 1560
DB 1501 XPDLPKTSKGVLELIPKVIHYKDLFPPTTSNGSPGHLDLVEGSLLOQTGEGAIKWEANRP 1560

QY 1561 GKVPPLRVATESSAKTPSKLLDPLAWDNHYGTQIPKEBWKSOEKSPKTAFFKKDTILSL 1620
DB 1561 GKVPPLRVATESSAKTPSKLLDPLAWDNHYGTQIPKEBWKSOEKSPKTAFFKKDTILSL 1620

QY 1621 NACESNHAIAAINEGONKPEIEVTWAKQRTERLCSQNPPLVKRQREBITTTLSQDQEE 1680
DB 1621 NACESNHAIAAINEGONKPEIEVTWAKQRTERLCSQNPPLVKRQREBITTTLSQDQEE 1680

QY 1681 IDYDDTISVEMKKEDFDIYDEENQSPRSFOKTRHYFIAAVERLWDYGMSSSPVLNRR 1740
DB 1681 IDYDDTISVEMKKEDFDIYDEENQSPRSFOKTRHYFIAAVERLWDYGMSSSPVLNRR 1740

QY 1741 AOSGSVPQPKVVFQEFDTGSGTQPLVYGEHNEHGLLGPYIRAEVEDINIMVTFNQASR 1800
DB 1741 AOSGSVPQPKVVFQEFDTGSGTQPLVYGEHNEHGLLGPYIRAEVEDINIMVTFNQASR 1800

QY 1801 PYSFYSSLIISYEEQORQGAEPKKNPKVNETKTYFWKQHHMAPTKDFCKAWAYFSDV 1860

```

Db 1801 PYSFYSSLLSYEDQFQGAEPKKNFKPNETKTYFWKVOHMAPTKDEDCRAWAFSDV 1860
QY 1861 DLEKDVHSLGLPGLLVCHTNTNTPAHGRQVTVQEFALPFTTFDEYKSWYFTENMERNCSA 1920
Db 1861 DLEKDVHSLGLPGLLVCHTNTNTPAHGRQVTVQEFALPFTTFDEYKSWYFTENMERNCSA 1920
QY 1921 PCNIQMEDPTFKENYRFAHNGVIMDTPLCLVMAQDQIRWYLLSGNSNENIHSIHFSGH 1980
Db 1921 PCNIQMEDPTFKENYRFAHNGVIMDTPLCLVMAQDQIRWYLLSGNSNENIHSIHFSGH 1980
QY 1981 VFTVRKKEYKXALYNLYGVEETVEMLPKAGIWRVECLIGEHLAGMSTLFLVYSNKC 2040
Db 1981 VFTVRKKEYKXALYNLYGVEETVEMLPKAGIWRVECLIGEHLAGMSTLFLVYSNKC 2040
QY 2041 QTPLGNASGHIRDFQITAGSQVQOWAPKLARLHYSGSINAWSTKEPFSWKVDLLAPMII 2100
Db 2041 QTPLGNASGHIRDFQITAGSQVQOWAPKLARLHYSGSINAWSTKEPFSWKVDLLAPMII 2100
QY 2101 HGIKTQARQKTSLSYISOFIIMYSLDGKKWQTYRGNSTGTLMVFFGNVDSSGIXHNIFN 2160
Db 2101 HGIKTQARQKTSLSYISOFIIMYSLDGKKWQTYRGNSTGTLMVFFGNVDSSGIXHNIFN 2160
QY 2161 PPIIARYIRLPHTHYSIRTLRMLMGCGLNCSMPLGMESKAISDAQITASSYFTNMPFA 2220
Db 2161 PPIIARYIRLPHTHYSIRTLRMLMGCGLNCSMPLGMESKAISDAQITASSYFTNMPFA 2220
QY 2221 TWSPSKARLHLOGRSNAMPQVNNPKEMIQVDFQKTMKVTVGTTQGVKSLLSMYKKEFL 2280
Db 2221 TWSPSKARLHLOGRSNAMPQVNNPKEMIQVDFQKTMKVTVGTTQGVKSLLSMYKKEFL 2280
QY 2281 ISSSQDGHQWTLFFQNGKVKVFGQNDSDPTFPVNS:DPPLLTRYLRIHPQSWVHQIALRM 2340
Db 2281 ISSSQDGHQWTLFFQNGKVKVFGQNDSDPTFPVNS:DPPLLTRYLRIHPQSWVHQIALRM 2340
QY 2341 EVLGCEAQDLY 2351
Db 2341 EVLGCEAQDLY 2351

```

RESULT 9

AAW11435
ID AAW11435 standard; protein; 2351 AA.

AC AAW11435;

DT 20-NOV-1997 (first entry)

DE Active Factor VIII:C analogue S1311X.

XX Factor VIII:C; analogue; glycoprotein; blood coagulation cascade;
KW fibrinogen; fibrin clot; haemostasis; haemophilia A; bleeding diathesis;
KW plasma protease; thrombin; immunogen; antibody; haemophilic; therapy;
XX proteolytic cleavage.

OS Homo sapiens.
OS Synthetic.

EH Key Location/Qualifiers

FT Peptide 1..19 /note= "signal peptide"

FT Protein 20..2351 /note= "mature Factor VIII:C"

FT Region 20..1667 /note= "heavy chain fragment"

FT Domain 760..1667 /note= "B domain"

FT Modified-site 1330

FT /label= Phe, Glu, Pro

FT Region 1668..2350

FT /note= "light chain fragment"

XX

XX

PN W09703195-A1.

XX 30-JAN-1997.

XX 09-JUL-1996; 96WO-US011444.

XX 11-JUL-1995; 95US-0001025P.

XX (CHIR) CHIRON CORP.

XX Hung DT, Cohen FE, Innis N;

XX WPI; 1997-119050/11.

XX Factor VIII:C analog modified adjacent to a non-activating Arg residue -
used in the treatment of haemophilias, by improvement of haemostasis.

XX Claim 30; Page; 90pp; English.

XX AAW11330-W11472 represent active Factor VIII:C analogues of the
invention. These sequences were created by mutating the wild type factor
VIII:C coding sequence (see AAT51357) using mutagenic primers. The
analogues comprise a native Factor VIII:C polypeptide modified at a site
adjacent to a non-activating Arg residue so that a Arg-Pro or Pro-Arg
dipeptide is created. Factor VIII:C is a large glycoprotein that
participates in the blood coagulation cascade that ultimately converts
soluble fibrinogen to insoluble fibrin clot, effecting haemostasis. A
deficiency in Factor VIII:C is responsible for haemophilia A, which is an
X-chromosome-linked inherited bleeding diathesis. Factor VIII:C is
activated by plasma proteases, such as thrombin. During activation the
mature polypeptide is cleaved to generate heavy and light chain fragments
that are further cleaved. Complexes of two or more of the analogues,
nucleic acids and vectors encoding them may be used alone or in
conjunction with each other, for the prevention or treatment of active
Factor VIII:C deficiency in a mammal. The analogues may be used as
immunogens to raise antibodies, and in the treatment of haemophilias, by
improvement of haemostasis. The analogues are resistant to proteolytic
cleavage and display increased plasma half-life. They may be administered
at lower dosages and by different modes of administration

XX Sequence 2351 AA;

Query Match 99.8%; Score 12395; DB 2; Length 2351;

Best Local Similarity 99.8%; Pred. No. 0;

Matches 2347; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 MOIELSTCFFCLLPFCFSATRRYYLGAVELSWDYQSDLGELPYDARPPRPVPSFFN 60

Db 1 MOIELSTCFFCLLPFCFSATRRYYLGAVELSWDYQSDLGELPYDARPPRPVPSFFN 60

QY 61 TSVVYKKTFLVEFTDHLFNIAKPRPPMGLLGPITQAEVYDTVTITLKNMASHPVSLHAV 120

Db 61 TSVVYKKTFLVEFTDHLFNIAKPRPPMGLLGPITQAEVYDTVTITLKNMASHPVSLHAV 120

QY 121 GVSYWKASEGAYDDQTSQREKEDDKVFPQSGSHTVVQVLKENGPMASDPLCLTYSYLSH 180

Db 121 GVSYWKASEGAYDDQTSQREKEDDKVFPQSGSHTVVQVLKENGPMASDPLCLTYSYLSH 180

QY 181 VDLVKDLSGLIGALLVCREGSLAKEKTQTLHKFTLLFAVDFEGKSWHSETKNSLMQDSD 240

Db 181 VDLVKDLSGLIGALLVCREGSLAKEKTQTLHKFTLLFAVDFEGKSWHSETKNSLMQDSD 240

QY 241 AASARAWPKMHTVNGYVNRSLPGLIGCHRKSVYWHVIGMTTTPPEVHSIFLEGHTFLVRNH 300

Db 241 AASARAWPKMHTVNGYVNRSLPGLIGCHRKSVYWHVIGMTTTPPEVHSIFLEGHTFLVRNH 300

QY 301 ROASLEISPIITFLTAQTLLMDLQGFLLSCHSHQHDGMEAYKVVDSCPEPQIMKQNE 360

Db 301 ROASLEISPIITFLTAQTLLMDLQGFLLSCHSHQHDGMEAYKVVDSCPEPQIMKQNE 360

QY 361 EAEDYDDDLTDSMDVVRFDNDSPTQIRSVAKKHPTKVHVYIAAEEEDWDYAPLVLA 420

Db 361 EAEDYDDDLTDSMDVVRFDNDSPTQIRSVAKKHPTKVHVYIAAEEEDWDYAPLVLA 420

QY 421 PDDRSYKQYLNNGQRIQKRYKVRFMAYTDETEKTRAIQHESGILGPLYGVGDTL 480
DB 421 PDDRSYKQYLNNGQRIQKRYKVRFMAYTDETEKTRAIQHESGILGPLYGVGDTL 480
QY 481 LIIFXQASRPNIYFPHGITHVRPLYSRRLPKGVHKLKDFPLPOBIYKRYKWTVTVEDGP 540
DB 481 LIIFXQASRPNIYFPHGITHVRPLYSRRLPKGVHKLKDFPLPOBIYKRYKWTVTVEDGP 540
QY 541 TKSDPRCLTRYYSFVNMRDGLASGLIGPLLICYKESVDKGNQMSDKRNVLFSVDE 600
DB 541 TKSDPRCLTRYYSFVNMRDGLASGLIGPLLICYKESVDKGNQMSDKRNVLFSVDE 600
QY 601 NRSWYLTENIQRLPNPAGVQLDEPEFQASNIHMSINGVYFDSLQSLVCLHEVAYWYILS 660
DB 601 NRSWYLTENIQRLPNPAGVQLDEPEFQASNIHMSINGVYFDSLQSLVCLHEVAYWYILS 660
QY 661 IGAQTDFLSVFSGYTFKHMYVEDTLFLPPSGGTHVPMENPGLWILGCHNSDFRNRG 720
DB 661 IGAQTDFLSVFSGYTFKHMYVEDTLFLPPSGGTHVPMENPGLWILGCHNSDFRNRG 720
QY 721 MTALLKVSSCDKNTGDYEDSDYEDISAYLLSXNAIEPRSFQNSRHHSTROKQENATTI 780
DB 721 MTALLKVSSCDKNTGDYEDSDYEDISAYLLSXNAIEPRSFQNSRHHSTROKQENATTI 780
QY 781 PENDIEKTPWFAHRTMPKIQNVSSDILMLRQSPHGLSLDLQEAKEYETSDDPSS 840
DB 781 PENDIEKTPWFAHRTMPKIQNVSSDILMLRQSPHGLSLDLQEAKEYETSDDPSS 840
QY 841 PCAIDSNLSBWHFRPOLHHSQDMVTPESGLQRLNEKLGTTAAATELKKLDFKVSST 900
DB 841 PCAIDSNLSBWHFRPOLHHSQDMVTPESGLQRLNEKLGTTAAATELKKLDFKVSST 900
QY 901 SNNLISTIPSONLAAGTONTSLGPPMPVHVDSDTLFLGKKSPLTESGGPLSLSEE 960
DB 901 SNNLISTIPSONLAAGTONTSLGPPMPVHVDSDTLFLGKKSPLTESGGPLSLSEE 960
QY 961 NNDKLLRGLMNSQESSWGKNVSTESGRLPKGRAGPALLTKDNALFKVSIILKTN 1020
DB 961 NNDKLLRGLMNSQESSWGKNVSTESGRLPKGRAGPALLTKDNALFKVSIILKTN 1020
QY 1021 KTSNNAATNRKTHIDGSPILLIENSVMONILLESDETEPKVTPLTHDRMLMDKNATLRL 1080
DB 1021 KTSNNAATNRKTHIDGSPILLIENSVMONILLESDETEPKVTPLTHDRMLMDKNATLRL 1080
QY 1081 NHMSNKTSSKNMNVQKKEGIPPDACNPDMSPFQMLFLPESARWQRTGKNSLNSG 1140
DB 1081 NHMSNKTSSKNMNVQKKEGIPPDACNPDMSPFQMLFLPESARWQRTGKNSLNSG 1140
QY 1141 QGPSKQLVSLGPEKSVQCNFLSEKNKVYVKGKPTKDVGLKEMVFPSSRNLFITNLN 1200
DB 1141 QGPSKQLVSLGPEKSVQCNFLSEKNKVYVKGKPTKDVGLKEMVFPSSRNLFITNLN 1200
QY 1201 LHENNTNHOEKI QBEIEKKEFTLIQENVVLPQIHTVTGTPKPMKNLFLLLSTRQNVGSYD 1260
DB 1201 LHENNTNHOEKI QBEIEKKEFTLIQENVVLPQIHTVTGTPKPMKNLFLLLSTRQNVGSYD 1260
QY 1261 GAYAVLQDFRSLNDSNTRKKTAAHFKKGBEENLEGNGTKQIVERYACTTRISPT 1320
DB 1261 GAYAVLQDFRSLNDSNTRKKTAAHFKKGBEENLEGNGTKQIVERYACTTRISPT 1320
QY 1321 SOQNVTORSEKALQKPLPLEETELEKRIIYDDTSTOWSKNMKHLTPSTLTQIDYNEKE 1380
DB 1321 SOQNVTORSEKALQKPLPLEETELEKRIIYDDTSTOWSKNMKHLTPSTLTQIDYNEKE 1380
QY 1381 KGAITQSPSLDCLTRSHSIPQANRPLPAKYSSFPSPRIYLTTRVLFDNSSHLPAAASY 1440
DB 1381 KGAITQSPSLDCLTRSHSIPQANRPLPAKYSSFPSPRIYLTTRVLFDNSSHLPAAASY 1440
QY 1441 RKKDSGVQESSHFLQCAKKNLSAILTLEMTGDQREVSGLSGTSATNSVYTKKVENTVLP 1500
DB 1441 RKKDSGVQESSHFLQCAKKNLSAILTLEMTGDQREVSGLSGTSATNSVYTKKVENTVLP 1500
QY 1501 KFDLPKTSKQVLELLPKVHIYOKDLFPFTETSNQSPGHLDLVEGSLIQTEGAIKWNEANRP 1560

DB 1501 KFDLPKTSKQVLELLPKVHIYOKDLFPFTETSNQSPGHLDLVEGSLIQTEGAIKWNEANRP 1560
QY 1561 GKVPPLRVATSSAKTPSKLLDPLAWDNHYGTQIPKEEWKQOEKSPKTAFFKKDTIISL 1620
DB 1561 GKVPPLRVATSSAKTPSKLLDPLAWDNHYGTQIPKEEWKQOEKSPKTAFFKKDTIISL 1620
QY 1621 NACRSNHAIAANEGONKPEIEVTWAKQRTERRCSONPPVLKGFORBITTTIQLSDOEE 1680
DB 1621 NACRSNHAIAANEGONKPEIEVTWAKQRTERRCSONPPVLKGFORBITTTIQLSDOEE 1680
QY 1681 IDYDDTISVENKKEDFDIYDEDENQSPRSFQKTRHYFIAAVERLWDYCMSSSPHVLNRR 1740
DB 1681 IDYDDTISVENKKEDFDIYDEDENQSPRSFQKTRHYFIAAVERLWDYCMSSSPHVLNRR 1740
QY 1741 AQSGSVPOFKKVPQCEFTDGSFTQPLIRGELENEHLLGLGPVIRARVEDNIMVTFNQASR 1800
DB 1741 AQSGSVPOFKKVPQCEFTDGSFTQPLIRGELENEHLLGLGPVIRARVEDNIMVTFNQASR 1800
QY 1801 PYSFYSSLSIYEEQROGAEPKPNKPVKNETKTYFWKQVQHMAPTKDRFDCKAWAYFSDV 1860
DB 1801 PYSFYSSLSIYEEQROGAEPKPNKPVKNETKTYFWKQVQHMAPTKDRFDCKAWAYFSDV 1860
QY 1861 DLEKDVHSLGILGPLLVCHTNTLNPAGHQVTVQBFALPFTTIDETKSMVFTENMERNCRA 1920
DB 1861 DLEKDVHSLGILGPLLVCHTNTLNPAGHQVTVQBFALPFTTIDETKSMVFTENMERNCRA 1920
QY 1921 PCNIOMEDPTKENYRPHAINGYINDTLPGLVMAQDQIRWYLLSMGSGNENIHSIHFSGH 1980
DB 1921 PCNIOMEDPTKENYRPHAINGYINDTLPGLVMAQDQIRWYLLSMGSGNENIHSIHFSGH 1980
QY 1981 VFTVRKKEEYKMALYNLYPGVFETVEMLPKAGIWRVECLIGEHLHAGMSTLFLVYSNKC 2040
DB 1981 VFTVRKKEEYKMALYNLYPGVFETVEMLPKAGIWRVECLIGEHLHAGMSTLFLVYSNKC 2040
QY 2041 QTPLGMAHGHIRDPOITASGOYGOWAPKLARLHYSGSNAMSTKEPFSWKVDLLAPMII 2100
DB 2041 QTPLGMAHGHIRDPOITASGOYGOWAPKLARLHYSGSNAMSTKEPFSWKVDLLAPMII 2100
QY 2101 HGIKTQAGARQFSSLIYSQFIIMYSLDGKKWQTYRGNSTGTLMVFFGNVDSSGIKHNIFN 2160
DB 2101 HGIKTQAGARQFSSLIYSQFIIMYSLDGKKWQTYRGNSTGTLMVFFGNVDSSGIKHNIFN 2160
QY 2161 PPIIARYIRLHPTFYSIRSTRMELMGCDLNSCMPLGWESKAIISDAQITASSYFTNMEA 2220
DB 2161 PPIIARYIRLHPTFYSIRSTRMELMGCDLNSCMPLGWESKAIISDAQITASSYFTNMEA 2220
QY 2221 TWSPSKARLHLQGRSNAWRPQVNNPKEWLQVDFQKTMKVTGVTGQVKSLLTSMYVKEFL 2280
DB 2221 TWSPSKARLHLQGRSNAWRPQVNNPKEWLQVDFQKTMKVTGVTGQVKSLLTSMYVKEFL 2280
QY 2281 ISSSDQGHQWTLFFONGKVKVQGNQDSFTPVVNSLDPLLTRYLRIHPQSVWVHOIALRM 2340
DB 2281 ISSSDQGHQWTLFFONGKVKVQGNQDSFTPVVNSLDPLLTRYLRIHPQSVWVHOIALRM 2340
QY 2341 EVLGCEAQDLY 2351
DB 2341 EVLGCEAQDLY 2351
RESULT 10
AAW11445
ID AAW11445 standard; protein; 2351 AA.
XX
AC AAW11445;
XX
DT 20-NOV-1997 (first entry)
XX
DE Active Factor VIII:C analogue L1643X.
XX
KW Factor VIII:C; analogue; glycoprotein; blood coagulation cascade;
KW fibrinogen; fibrin clot; haemostasis; haemophilia A; bleeding diathesis;
KW plasma protease; thrombin; immunogen; antibody; haemophilic; therapy;

proteolytic cleavage.

Homo sapiens.
Synthetic.

Key Location/Qualifiers
1..19
/note= "signal peptide"
Protein 20..2351
/note= "mature Factor VIII:C"
Region 20..1667
/note= "heavy chain fragment"
Domain 760..1667
/note= "B domain"
Modified-site 1662
/label= Phe, Glu, Pro
Region 1668..2350
/note= "light chain fragment"
WO9703195-A1.
30-JAN-1997.
09-JUL-1996; 95WO-US011444.
11-JUL-1995; 95US-0001025P.
(CHIR) CHIRON CORP.
Hung DT, Cohen FE, Innis M;
WPI; 1997-119050/11.
Factor VIII:C analog modified adjacent to a non-activating Arg residue -
used in the treatment of haemophiliacs, by improvement of haemostasis.
Claim 32; Page; 90pp; English.
AAW11330-W11472 represent active Factor VIII:C analogues of the
invention. These sequences were created by mutating the wild type Factor
VIII:C coding sequence (see AAW51357) using mutagenic primers. The
analogues comprise a native Factor VIII:C polypeptide modified at a site
adjacent to a non-activating Arg residue so that a Arg-Pro or Pro-Arg
dipeptide is created. Factor VIII:C is a large glycoprotein that
participates in the blood coagulation cascade that ultimately converts
soluble fibrinogen to insoluble fibrin clot, effecting haemostasis. A
deficiency in Factor VIII:C is responsible for haemophilia A, which is an
X-chromosome-linked inherited bleeding diathesis. Factor VIII:C is
activated by plasma proteases, such as thrombin. During activation the
mature polypeptide is cleaved to generate heavy and light chain fragments
that are further cleaved. Complexes of two or more of the analogues,
nucleic acids and vectors encoding them may be used alone or in
conjunction with each other, for the prevention or treatment of active
Factor VIII:C deficiency in a mammal. The analogues may be used as
immunogens to raise antibodies, and in the treatment of haemophiliacs, by
improvement of haemostasis. The analogues are resistant to proteolytic
cleavage and display increased plasma half-life. They may be administered
at lower dosages and by different modes of administration
Sequence 2351 AA;

Query Match 99.8%; Score 12394; DB 2; Length 2351;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 2347; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

1 XQIELSTCFPLCLRRCFATRIYYLIGAVELSDVNMQSDLGELPYDARPPRPVPSFPPN 60
1 XQIELSTCFPLCLRRCFATRIYYLIGAVELSDVNMQSDLGELPYDARPPRPVPSFPPN 60
61 TSVYKKTFLVETFDLHFNIAKPRPPPMGLGPTIOAEVYDVTWVPLKNMASHPVSLHAV 120
61 TSVYKKTFLVETFDLHFNIAKPRPPPMGLGPTIOAEVYDVTWVPLKNMASHPVSLHAV 120

QY 121 GVSVMKASEGAEYDDQTSOREKEDDKVPFGSGSHYVWQVLKENGPMASDPCLCLTVSYLSH 180
DB 121 GVSVMKASEGAEYDDQTSOREKEDDKVPFGSGSHYVWQVLKENGPMASDPCLCLTVSYLSH 180
QY 181 VDLVKDLSNGLIGALLVCRSGSLAKEKTQTHKFIILLPAVFDGKSHSETKNSIMQORD 240
DB 181 VDLVKDLSNGLIGALLVCRSGSLAKEKTQTHKFIILLPAVFDGKSHSETKNSIMQORD 240
QY 241 AASARAWPMHTVNGYVNRSLPGLIGCHRKSVVHVIWGMGTTPVHSTFLGSHFLVFNH 300
DB 241 AASARAWPMHTVNGYVNRSLPGLIGCHRKSVVHVIWGMGTTPVHSTFLGSHFLVFNH 300
QY 301 RQASLEISPIITFLTAQTLLMDLGOFLLSCHISSHQHDGMEAYVKVDSCPEPQIMKNE 360
DB 301 RQASLEISPIITFLTAQTLLMDLGOFLLSCHISSHQHDGMEAYVKVDSCPEPQIMKNE 360
QY 361 EAEDYDDDLTSEMDVTRFDDDNSPFIQIRSVAKKHPTWVHYIAAEBEWDVAPLVIA 420
DB 361 EAEDYDDDLTSEMDVTRFDDDNSPFIQIRSVAKKHPTWVHYIAAEBEWDVAPLVIA 420
QY 421 PDDRSYKSOYLNNGPQIRGRYKVRPMAYTDEFTKREAIQHESGILGPLYGEGVDTL 480
DB 421 PDDRSYKSOYLNNGPQIRGRYKVRPMAYTDEFTKREAIQHESGILGPLYGEGVDTL 480
QY 481 LIIFKQASRPYNITYPHGITDVRPLYSRRLPKGVKHLKDFPILPGEIFKYKWTVTVEDGP 540
DB 481 LIIFKQASRPYNITYPHGITDVRPLYSRRLPKGVKHLKDFPILPGEIFKYKWTVTVEDGP 540
QY 541 TKSDPRCLTRYSSFVNMERDLASGLIGPLLCYKESVDQGNQIMSKRVILFSVDE 600
DB 541 TKSDPRCLTRYSSFVNMERDLASGLIGPLLCYKESVDQGNQIMSKRVILFSVDE 600
QY 601 NRSWLTENIORFLPNPAGVOLEDPEFOASNMHSINGYVFDLSOLSVCLEHVAYWYLS 660
DB 601 NRSWLTENIORFLPNPAGVOLEDPEFOASNMHSINGYVFDLSOLSVCLEHVAYWYLS 660
QY 661 TGAQDFLSVFGSYTFKHKVYEDTLTLPFSGETVFMENPGLMILGCHNSDFNRNG 720
DB 661 TGAQDFLSVFGSYTFKHKVYEDTLTLPFSGETVFMENPGLMILGCHNSDFNRNG 720
QY 721 WTALLKVSCKNTGDIYEDSDVEDISAYLLSKNNAIEPRSPQSRHSPSTQKQFNATTI 780
DB 721 WTALLKVSCKNTGDIYEDSDVEDISAYLLSKNNAIEPRSPQSRHSPSTQKQFNATTI 780
QY 781 PENDIEKTDPAFAHRTPMPTKIQNVSSSLLMLLRQSPFHGILSLDIQEAKEYTFSDPS 840
DB 781 PENDIEKTDPAFAHRTPMPTKIQNVSSSLLMLLRQSPFHGILSLDIQEAKEYTFSDPS 840
QY 841 PGADSNNSLSEMTFRPOLHSHGDMVTPESGLQLRNEKLGTTAATLKKLDFKVSST 900
DB 841 PGADSNNSLSEMTFRPOLHSHGDMVTPESGLQLRNEKLGTTAATLKKLDFKVSST 900
QY 901 SNMLSTIPSDMLAAGTNTSSLGPPSMVHYDSQDLDITLFGKSSPLTESGGPLSLSEE 960
DB 901 SNMLSTIPSDMLAAGTNTSSLGPPSMVHYDSQDLDITLFGKSSPLTESGGPLSLSEE 960
QY 961 NNDSKLLBSGLMNSQESSWGKNVSTESGRLFKGRHAGPALLTKDNALFKVYSILKTN 1020
DB 961 NNDSKLLBSGLMNSQESSWGKNVSTESGRLFKGRHAGPALLTKDNALFKVYSILKTN 1020
QY 1021 KTSNNSATNRKTHIDGPSLLIENSFVQNILESDETEFKVTPLIHDRMLDKNATLRL 1080
DB 1021 KTSNNSATNRKTHIDGPSLLIENSFVQNILESDETEFKVTPLIHDRMLDKNATLRL 1080
QY 1081 NHMSKNTSSKNMVMQCKEGPIPPDAQNPDMSFPKMLFLPESARWIORHGNKNSNG 1140
DB 1081 NHMSKNTSSKNMVMQCKEGPIPPDAQNPDMSFPKMLFLPESARWIORHGNKNSNG 1140
QY 1141 QGFSFKQLVSLGPEKSVBQNFLEKKNVAVGKGEFTKDVGLKEMVFPSSRNLFITLNDN 1200
DB 1141 QGFSFKQLVSLGPEKSVBQNFLEKKNVAVGKGEFTKDVGLKEMVFPSSRNLFITLNDN 1200
QY 1201 LHENNTNQEKKIQEIEKKTETLIQENVVLVQJHTVTGTQKNFMKNLFLLLSTRQNEGSYD 1260

Db 1201 LHENNTNQEKI QEEIEKKE*LIQENVVLPQIHVTVG*TKNFMKNL*FLLST*QNVEGSD 1260
QY 1261 GAYAVLQDPSLNDSTNR*TKGHTAHFSKKGEE*NLGEG*QTKQI*VEKYACTTRIS*NT 1320
Db 1261 GAYAVLQDPSLNDSTNR*TKGHTAHFSKKGEE*NLGEG*QTKQI*VEKYACTTRIS*NT 1320
QY 1321 SQONFVTVQSKALQKFR*LPLEET*BLEKRI*IVDDT*STOWSKNMKHL*TPSTLT*QIDYNEKE 1380
Db 1321 SQONFVTVQSKALQKFR*LPLEET*BLEKRI*IVDDT*STOWSKNMKHL*TPSTLT*QIDYNEKE 1380
QY 1381 KGAIQTSPSLDC*LTSHS*IPQANRSP*LP*IAKVSSP*PSIRPI*YLT*RVLFQDNSS*HLP*AA*SY 1440
Db 1381 KGAIQTSPSLDC*LTSHS*IPQANRSP*LP*IAKVSSP*PSIRPI*YLT*RVLFQDNSS*HLP*AA*SY 1440
QY 1441 RKDQSGVQSS*HFLQAGAKNN*LS*LA*TL*EMT*GDQREV*GSL*GTS*ATNS*VY*KK*VNT*VLP 1500
Db 1441 RKDQSGVQSS*HFLQAGAKNN*LS*LA*TL*EMT*GDQREV*GSL*GTS*ATNS*VY*KK*VNT*VLP 1500
QY 1501 KPDLPTSGV*ELL*PKVHI*YQKDL*PPT*ET*SN*GSP*GHLD*VEGSL*LG*GT*EG*ALK*WNE*ANRP 1560
Db 1501 KPDLPTSGV*ELL*PKVHI*YQKDL*PPT*ET*SN*GSP*GHLD*VEGSL*LG*GT*EG*ALK*WNE*ANRP 1560
QY 1561 GKVPFLRVAT*ESSAKT*PSKLLD*PLAWDN*HYGT*QI*PK*EWK*SQ*EK*SP*EKT*AF*KK*OT*IL*SL 1620
Db 1561 GKVPFLRVAT*ESSAKT*PSKLLD*PLAWDN*HYGT*QI*PK*EWK*SQ*EK*SP*EKT*AF*KK*OT*IL*SL 1620
QY 1621 NACENH*AL*AA*INEG*ON*KE*IE*V*WAK*Q*G*TR*EL*CS*ON*PP*VL*KR*Q*RE*IT*TR*TL*OS*DO*EE 1680
Db 1621 NACENH*AL*AA*INEG*ON*KE*IE*V*WAK*Q*G*TR*EL*CS*ON*PP*VL*KR*Q*RE*IT*TR*TL*OS*DO*EE 1680
QY 1681 IDYDDT*IS*VEM*KK*ED*FD*IV*DE*NO*SP*RS*FQ*KK*TR*HY*F*IA*V*ER*L*MD*Y*G*SS*P*H*V*L*N*R 1740
Db 1681 IDYDDT*IS*VEM*KK*ED*FD*IV*DE*NO*SP*RS*FQ*KK*TR*HY*F*IA*V*ER*L*MD*Y*G*SS*P*H*V*L*N*R 1740
QY 1741 AQSGSV*PQ*KKV*V*Q*E*FT*DG*ST*Q*P*LY*RG*EL*NE*HL*GL*GP*Y*IR*AE*VED*NI*MT*FP*NN*Q*AS*R 1800
Db 1741 AQSGSV*PQ*KKV*V*Q*E*FT*DG*ST*Q*P*LY*RG*EL*NE*HL*GL*GP*Y*IR*AE*VED*NI*MT*FP*NN*Q*AS*R 1800
QY 1801 PYSF*YSS*LL*IS*Y*ED*OR*Q*G*RE*PK*NF*VK*NET*Y*F*WK*Q*H*MA*P*TK*DE*P*CK*AW*AY*F*SD*V 1860
Db 1801 PYSF*YSS*LL*IS*Y*ED*OR*Q*G*RE*PK*NF*VK*NET*Y*F*WK*Q*H*MA*P*TK*DE*P*CK*AW*AY*F*SD*V 1860
QY 1861 DLEKDV*H*SL*GL*PL*V*CH*VT*NT*LN*PA*H*GR*Q*V*TV*Q*E*F*AL*P*TI*F*DE*TK*SW*Y*FT*EN*ME*RN*CR*A 1920
Db 1861 DLEKDV*H*SL*GL*PL*V*CH*VT*NT*LN*PA*H*GR*Q*V*TV*Q*E*F*AL*P*TI*F*DE*TK*SW*Y*FT*EN*ME*RN*CR*A 1920
QY 1921 PCNI*Q*ME*DP*TK*EN*YR*F*H*AI*NG*Y*IM*DT*LP*GL*Y*MA*Q*DO*RI*W*Y*LL*SM*GN*EN*IH*SI*H*F*SG*H 1980
Db 1921 PCNI*Q*ME*DP*TK*EN*YR*F*H*AI*NG*Y*IM*DT*LP*GL*Y*MA*Q*DO*RI*W*Y*LL*SM*GN*EN*IH*SI*H*F*SG*H 1980
QY 1981 VFTVR*KK*E*Y*K*MA*LY*NP*GV*P*ET*V*EM*LP*SK*AG*IR*VE*CL*IG*EH*H*AG*W*ST*FL*V*Y*SN*CK 2040
Db 1981 VFTVR*KK*E*Y*K*MA*LY*NP*GV*P*ET*V*EM*LP*SK*AG*IR*VE*CL*IG*EH*H*AG*W*ST*FL*V*Y*SN*CK 2040
QY 2041 QTP*LG*W*AG*SH*IR*DF*Q*IT*AS*G*Q*G*W*AP*KL*AR*LY*H*Y*GS*IN*AW*ST*KB*PP*SM*K*IV*DL*LA*P*MI 2100
Db 2041 QTP*LG*W*AG*SH*IR*DF*Q*IT*AS*G*Q*G*W*AP*KL*AR*LY*H*Y*GS*IN*AW*ST*KB*PP*SM*K*IV*DL*LA*P*MI 2100
QY 2101 HG*IK*TO*GA*RO*Q*ESS*LY*IS*Q*FI*IM*Y*SL*DG*KK*W*O*Y*G*NS*ST*GT*LM*V*P*GN*V*DS*GI*KH*EN*IF*N 2160
Db 2101 HG*IK*TO*GA*RO*Q*ESS*LY*IS*Q*FI*IM*Y*SL*DG*KK*W*O*Y*G*NS*ST*GT*LM*V*P*GN*V*DS*GI*KH*EN*IF*N 2160
QY 2161 P*II*ARY*IR*H*E*TH*YS*IR*ST*LR*EL*MG*CD*LN*CS*PL*G*ES*KA*IS*DA*Q*IT*AS*SY*FT*NN*F*A 2220
Db 2161 P*II*ARY*IR*H*E*TH*YS*IR*ST*LR*EL*MG*CD*LN*CS*PL*G*ES*KA*IS*DA*Q*IT*AS*SY*FT*NN*F*A 2220
QY 2221 TW*SP*SK*AR*H*LG*QR*NA*RP*Q*VN*PK*EM*L*Q*VD*F*Q*TK*MY*G*VT*TO*GV*Z*SL*TS*MY*V*KE*FL 2280
Db 2221 TW*SP*SK*AR*H*LG*QR*NA*RP*Q*VN*PK*EM*L*Q*VD*F*Q*TK*MY*G*VT*TO*GV*Z*SL*TS*MY*V*KE*FL 2280
QY 2281 ISSSQD*GH*W*TL*FP*Q*G*K*V*F*Q*G*Q*DS*FT*P*V*NS*LD*P*LL*TR*Y*LR*IH*P*Q*SW*H*Q*IA*LR*M 2340

Db 2281 ISSSQD*GH*W*TL*FP*Q*G*K*V*F*Q*G*Q*DS*FT*P*V*NS*LD*P*LL*TR*Y*LR*IH*P*Q*SW*H*Q*IA*LR*M 2340
QY 2341 EV*LG*CE*AO*DL*Y 2351
Db 2341 EV*LG*CE*AO*DL*Y 2351
RESULT 11
AAW11398
ID AAW11398 standard; protein; 2351 AA.
XX AAW11398;
XX 18-NOV-1997 (first entry)
DE Active Factor VIII:C analogue, delta 746, + residue 746 insertion.
XX
KW Factor VIII:C; analogue; glycoprotein; blood coagulation cascade;
KW fibrinogen; fibrin clot; haemostasis; haemophilia A; bleeding diathesis;
KW plasma protease; thrombin; immunogen; antibody; haemophilic; therapy;
KW proteolytic cleavage.
XX Homo sapiens.
CS Synthetic.
XX
FH Key
FT Peptide
FT Protein
FT Region
FT Domain
FT Misc-difference
FT Modified-site
FT Region
FT Region
XX WO9703195-A1.
XX 30-JAN-1997.
XX 09-JUL-1996; 96WO-US011444.
XX 11-JUL-1995; 95US-0001025P.
XX (CHIR) CHIRON CORP.
XX Hung DT, Cohen FE, Innis X;
XX WPI; 1997-119050/11.
XX Factor VIII:C analog modified adjacent to a non-activating Arg residue -
XX used in the treatment of haemophilias, by improvement of haemostasis.
XX Claim 23; Page; 90pp; English.
XX AAW11398-W1472 represent active Factor VIII:C analogues of the
XX invention. These sequences were created by mutating the wild type Factor
XX VIII:C coding sequence (see AAT51357) using mutagenic primers. The
XX analogues comprise a native Factor VIII:C polypeptide modified at a site
XX adjacent to a non-activating Arg residue so that a Arg-Pro or Pro-Arg
XX dipeptide is created. Factor VIII:C is a large glycoprotein that
XX participates in the blood coagulation cascade that ultimately converts
XX soluble fibrinogen to insoluble fibrin clot, effecting haemostasis. A
XX deficiency in Factor VIII:C is responsible for haemophilia A, which is an
XX X-chromosome-linked inherited bleeding diathesis. Factor VIII:C is
XX activated by plasma proteases, such as thrombin. During activation the
XX mature polypeptide is cleaved to generate heavy and light chain fragments
XX that are further cleaved. Complexes of two or more of the analogues,

CC nucleic acids and vectors encoding them may be used alone or in
CC conjunction with each other, for the prevention or treatment of active
CC Factor VIII:C deficiency in a mammal. The analogues may be used as
CC immunogens to raise antibodies, and in the treatment of haemophiliacs, by
CC improvement of haemostasis. The analogues are resistant to proteolytic
CC cleavage and display increased plasma half-life. They may be administered
CC at lower dosages and by different modes of administration
XX
SQ Sequence 2351 AA;

Query Match 99.8%; Score 12394; DB 2; Length 2351;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 2347; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY	1	MQELSTCFCLLRFCFSATRRYYLGAVELSWDMQSD-GEIPVDAZFPPRPVKSPFFN	60
DB	1	MQ-ELSTCFCLLRFCFSATRRYYLGAVELSWDMQSD-GEIPVDAZFPPRPVKSPFFN	60
QY	61	TSVVYKKTLLVEFTDHLFNIAKRPFPWMLGPTIQAEVDTVVITLKNWASHPVSLHAV	120
DB	61	TSVVYKKTLLVEFTDHLFNIAKRPFPWMLGPTIQAEVDTVVITLKNWASHPVSLHAV	120
QY	121	GVSYWKASGAEXDDQTSOREKEDDKVPFGSSHTYVWVLKENGCPMASDPLCTYSYLSH	180
DB	121	GVSYWKASGAEXDDQTSOREKEDDKVPFGSSHTYVWVLKENGCPMASDPLCTYSYLSH	180
QY	181	VDLVKDLNSGLIGALLVCRGSLAKEXTQTLHKFILLFAVDFDEKSWHSETKSLMODRD	240
DB	181	VDLVKDLNSGLIGALLVCRGSLAKEXTQTLHKFILLFAVDFDEKSWHSETKSLMODRD	240
QY	241	AASARAWPKHTVNGYVNRSLPGLIGHRKSVYVHWVIGMTTPEVHSIFLEGHTFLVRNH	300
DB	241	AASARAWPKHTVNGYVNRSLPGLIGHRKSVYVHWVIGMTTPEVHSIFLEGHTFLVRNH	300
QY	301	RQASLETSPIITELTAOTLLMDLGQFLLSCHISHOHDGMEAYVKVDSCEPQIMKNNE	360
DB	301	RQASLETSPIITELTAOTLLMDLGQFLLSCHISHOHDGMEAYVKVDSCEPQIMKNNE	360
QY	361	EADYDDDLTDSMDVVRPDDNSPSFIQIRSAKHPKTHVHYIAAEEDWYAPLVLA	420
DB	361	EADYDDDLTDSMDVVRPDDNSPSFIQIRSAKHPKTHVHYIAAEEDWYAPLVLA	420
QY	421	PDRSYKSOVLNNGPQIRGRKYKVRPMAYTDETFKTRBAIQHSGILGELLGVEGVDTL	480
DB	421	PDRSYKSOVLNNGPQIRGRKYKVRPMAYTDETFKTRBAIQHSGILGELLGVEGVDTL	480
QY	481	LIIFKQASRPYNTYPHGITDVRPLYSRRLPKGVKHLKDPPILPGETFKYKWTVVEDGP	540
DB	481	LIIFKQASRPYNTYPHGITDVRPLYSRRLPKGVKHLKDPPILPGETFKYKWTVVEDGP	540
QY	541	TKSDPRLCTIYYSFVNMRDLASGLIGPLLI CYKESVDQKGNQIMSKDNVILFSVFDE	600
DB	541	TKSDPRLCTIYYSFVNMRDLASGLIGPLLI CYKESVDQKGNQIMSKDNVILFSVFDE	600
QY	601	NRSWYLTENIQRFLPNPAGVQLDPEPQASNIHMSINGYVDSIQSVCLHVAAYKYL	660
DB	601	NRSWYLTENIQRFLPNPAGVQLDPEPQASNIHMSINGYVDSIQSVCLHVAAYKYL	660
QY	661	IGAQTDFLSVFFSGYTFKRWVYEDTLTLPFGSETVFMSENENPGLWILGCHNSDFNRG	720
DB	661	IGAQTDFLSVFFSGYTFKRWVYEDTLTLPFGSETVFMSENENPGLWILGCHNSDFNRG	720
QY	721	YTAALLKVSXCDKNTGDIYSDYSEDISAYLLSKNAIEPRFSQNSRHPSTRQKQFNATTI	780
DB	721	YTAALLKVSXCDKNTGDIYSDYSEDISAYLLSKNAIEPRFSQNSRHPSTRQKQFNATTI	780
QY	781	PENDIEKTDWPAHRTPMKIQNVSSDLMLVLSQSTPHGLSLDQEAKEYTSDDPS	840
DB	781	PENDIEKTDWPAHRTPMKIQNVSSDLMLVLSQSTPHGLSLDQEAKEYTSDDPS	840
QY	841	PGAIDSNNSLSEWTHFRPQLHSGDMVFTPESGQLRLNEKLGTTAATLKKLDKFKVSS	900
DB	841	PGAIDSNNSLSEWTHFRPQLHSGDMVFTPESGQLRLNEKLGTTAATLKKLDKFKVSS	900

QY	901	SNNLISTIPSDNLAACTDNTSSIGPPSPVHYDSQDITLFGKKSPLTSGGPILSSE	960
DB	901	SNNLISTIPSDNLAACTDNTSSIGPPSPVHYDSQDITLFGKKSPLTSGGPILSSE	960
QY	961	NNDKLLSGLMNSQSSWGKNSVSTSGRLFKGKRAHGPALLTKONALFKVSIILKTN	1020
DB	961	NNDKLLSGLMNSQSSWGKNSVSTSGRLFKGKRAHGPALLTKONALFKVSIILKTN	1020
QY	1021	KTSNNSATNEKTHIDGPSILLIENSPSWQNILESDETFKKVPLIHDRMLMDKNAALRL	1080
DB	1021	KTSNNSATNEKTHIDGPSILLIENSPSWQNILESDETFKKVPLIHDRMLMDKNAALRL	1080
QY	1081	NHMSNKTTSKKNMEMVQOKKEGPIPPDAONPDMSPFKMLFLPESARWIOETHCKNSLNSG	1140
DB	1081	NHMSNKTTSKKNMEMVQOKKEGPIPPDAONPDMSPFKMLFLPESARWIOETHCKNSLNSG	1140
QY	1141	QGPSKQOLVSLGPEKVEGQNFLESEXKVVVGKGFVKDGLKEMVFPSSRNLFNLND	1200
DB	1141	QGPSKQOLVSLGPEKVEGQNFLESEXKVVVGKGFVKDGLKEMVFPSSRNLFNLND	1200
QY	1201	LHENNTHNOEKKLOEBIEYKETHIOENVVLPOIHTVGTGNFMKNLFLISTRQNEGSVD	1260
DB	1201	LHENNTHNOEKKLOEBIEYKETHIOENVVLPOIHTVGTGNFMKNLFLISTRQNEGSVD	1260
QY	1261	GAYAPVLQDFRSLNDSTNETKCHTAHFKKGBEENLEGLCNQTKQIVKVCATTRISPT	1320
DB	1261	GAYAPVLQDFRSLNDSTNETKCHTAHFKKGBEENLEGLCNQTKQIVKVCATTRISPT	1320
QY	1321	SOQFVTVQSKRALKQFRPLBEETELEKRIIVDDTSTQMSKNMKHLTPSTLQIDYNEKE	1380
DB	1321	SOQFVTVQSKRALKQFRPLBEETELEKRIIVDDTSTQMSKNMKHLTPSTLQIDYNEKE	1380
QY	1381	KGAITQSPSDCLTRSHSIPQANRSPPIAKYSSFPSIRPIYLYTRVLFODNSHLPAASY	1440
DB	1381	KGAITQSPSDCLTRSHSIPQANRSPPIAKYSSFPSIRPIYLYTRVLFODNSHLPAASY	1440
QY	1441	RKKGSGVQSSSHFLQAKAKNNLSLAILTLEMTGDQREVSGLTSATNSVYKKEVNTVLP	1500
DB	1441	RKKGSGVQSSSHFLQAKAKNNLSLAILTLEMTGDQREVSGLTSATNSVYKKEVNTVLP	1500
QY	1501	KDPLPTSKQVHELLPKVHIYQKDLPTFTSNGSPGHLDIVEGSLLOGTGAIKWNEANRP	1560
DB	1501	KDPLPTSKQVHELLPKVHIYQKDLPTFTSNGSPGHLDIVEGSLLOGTGAIKWNEANRP	1560
QY	1561	GKVPFLRVATESSAKTPSKLLDPLAWNHYGTQIPKEEWKSQKSEKFAFKKDTILSL	1620
DB	1561	GKVPFLRVATESSAKTPSKLLDPLAWNHYGTQIPKEEWKSQKSEKFAFKKDTILSL	1620
QY	1621	NACESNEAJAINEGQNKPEIEVTWAKQGETERLCSQNPPLVKRHOEITRRTTLQSDQEE	1680
DB	1621	NACESNEAJAINEGQNKPEIEVTWAKQGETERLCSQNPPLVKRHOEITRRTTLQSDQEE	1680
QY	1681	IDYDDTISVEMKKEDEPDYDEDEENOSPRFOKTRHYFIAAVERLWDYGMSSSPHVLNR	1740
DB	1681	IDYDDTISVEMKKEDEPDYDEDEENOSPRFOKTRHYFIAAVERLWDYGMSSSPHVLNR	1740
QY	1741	AQSGSVPPQKVVQFQFTDGSFTQPLRYGELNEHGLLGPYIRAEVEDNMVTFRNOASR	1800
DB	1741	AQSGSVPPQKVVQFQFTDGSFTQPLRYGELNEHGLLGPYIRAEVEDNMVTFRNOASR	1800
QY	1801	PYSFYSSLISYEDQROGAEPKRNFKVKNETKTYFWKQVHMAPTKDEDFCKAWAYFSV	1860
DB	1801	PYSFYSSLISYEDQROGAEPKRNFKVKNETKTYFWKQVHMAPTKDEDFCKAWAYFSV	1860
QY	1861	DLEKOVHSLGILGPLLVCHTNTINPAHGRQVTVQBFALFTTIFDETKSWYFTENMERNCA	1920
DB	1861	DLEKOVHSLGILGPLLVCHTNTINPAHGRQVTVQBFALFTTIFDETKSWYFTENMERNCA	1920
QY	1921	PCNIQWEDTFKKNYRFFAHNGYIMDTLPLGVMAQORIRWYLLSNGSNENIHSIHFSGH	1980
DB	1921	PCNIQWEDTFKKNYRFFAHNGYIMDTLPLGVMAQORIRWYLLSNGSNENIHSIHFSGH	1980

QY 601 NRSWYLTENIQRILPNPAGVQLEDPEPQASNMHSINGYVFDLSQLSVCLHEVAYWYLS 660
DB 601 NRSWYLTENIQRILPNPAGVQLEDPEPQASNMHSINGYVFDLSQLSVCLHEVAYWYLS 660
QY 661 IGAQTDELVSFFSGYTPKHQWVEDTTLTPFSGETVFMENPGLWILCHNSDFNRG 720
DB 661 IGAQTDELVSFFSGYTPKHQWVEDTTLTPFSGETVFMENPGLWILCHNSDFNRG 720
QY 721 MTALLKVSSCDKNTGDYEDSYEDISAYLLSKNAIEPRFSQNSHPSTQCFNATII 780
DB 721 MTALLKVSSCDKNTGDYEDSYEDISAYLLSKNAIEPRFSQNSHPSTQCFNATII 780
QY 781 PENDIEKTDWFAHRTMPKIQNVSSDDLMLLRQSPTHGLSLSDLOEAKYTFSDPS 840
DB 781 PENDIEKTDWFAHRTMPKIQNVSSDDLMLLRQSPTHGLSLSDLOEAKYTFSDPS 840
QY 841 PGADSNNSISEMTHPRQLHHSQDMVFTPEBSGLQLRLNEKLGHTAATELKKLDFKVSST 900
DB 841 PGADSNNSISEMTHPRQLHHSQDMVFTPEBSGLQLRLNEKLGHTAATELKKLDFKVSST 900
QY 901 SNNLISIPSDNLAAGTNTSSLGPPMPVHYDSQDITLFGKSSPLTSSGGLSLSEE 960
DB 901 SNNLISIPSDNLAAGTNTSSLGPPMPVHYDSQDITLFGKSSPLTSSGGLSLSEE 960
QY 961 NNDSKLLESGLMNSQSSWGKNSVSTESCLRFKRAEGPALLTKONALFKVSIULKTN 1020
DB 961 NNDSKLLESGLMNSQSSWGKNSVSTESCLRFKRAEGPALLTKONALFKVSIULKTN 1020
QY 1021 KTSNNSATNKTILDGPSLLIENSPPVQWNILEDSTEFKKVTLIHDRMLMDKNATLRL 1080
DB 1021 KTSNNSATNKTILDGPSLLIENSPPVQWNILEDSTEFKKVTLIHDRMLMDKNATLRL 1080
QY 1081 NMSNKTTSKKNMVMQKKEGIPDAQNPDVXFFKMLFLPESARWIQTHGKNSLNSG 1140
DB 1081 NMSNKTTSKKNMVMQKKEGIPDAQNPDVXFFKMLFLPESARWIQTHGKNSLNSG 1140
QY 1141 QGSPKQLVSLGPEKSVGQNFISEKNKVVGKGEFTKDVGLKEMVFPSSRNFLNLDN 1200
DB 1141 QGSPKQLVSLGPEKSVGQNFISEKNKVVGKGEFTKDVGLKEMVFPSSRNFLNLDN 1200
QY 1201 LHNNTNHOEKKIQEETEKETLIIQENVLPQHTVTKGNFMKNFLILSTRQWVGSYD 1260
DB 1201 LHNNTNHOEKKIQEETEKETLIIQENVLPQHTVTKGNFMKNFLILSTRQWVGSYD 1260
QY 1261 GAYAPVLQDFRSLNDSNTRKCHTAHFSKGBEENLEGLNQTKQIVEKVACTTRISPT 1320
DB 1261 GAYAPVLQDFRSLNDSNTRKCHTAHFSKGBEENLEGLNQTKQIVEKVACTTRISPT 1320
QY 1321 SQONFYTORSKRALKOPRLPLEBETELEKRIIVDDTSTQWSKNMHLTPSTLTQIDYNEXE 1380
DB 1321 SQONFYTORSKRALKOPRLPLEBETELEKRIIVDDTSTQWSKNMHLTPSTLTQIDYNEXE 1380
QY 1381 KGAITQSPISDCILTRSHSIPQANRSPLPYAKVSSFPISIRPIYLTVLFODNSHLEPAASY 1440
DB 1381 KGAITQSPISDCILTRSHSIPQANRSPLPYAKVSSFPISIRPIYLTVLFODNSHLEPAASY 1440
QY 1441 RKKDSGVQSSHPLOQAKXNNLSAILTLTMTGDQREVSLGTSATNSVTYKKVENTVLP 1500
DB 1441 RKKDSGVQSSHPLOQAKXNNLSAILTLTMTGDQREVSLGTSATNSVTYKKVENTVLP 1500
QY 1501 KPDLPKTSQKVELLPKVHIYQKDLFPTETSNQSGFHLDLVEGSLLOGTGAIKWNENRPP 1560
DB 1501 KPDLPKTSQKVELLPKVHIYQKDLFPTETSNQSGFHLDLVEGSLLOGTGAIKWNENRPP 1560
QY 1561 GKVPFLRVATESAKTPSKLLDPLANDNHYGTQIPKEWKSQKSPKTAFFKKOTILSL 1620
DB 1561 GKVPFLRVATESAKTPSKLLDPLANDNHYGTQIPKEWKSQKSPKTAFFKKOTILSL 1620
QY 1621 NACESNHAIAINEGQNKPEIEVTWAKQKTERPLCSQNPVPLKRRHOREITRITLQSDQEE 1680
DB 1621 NACESNHAIAINEGQNKPEIEVTWAKQKTERPLCSQNPVPLKRRHOREITRITLQSDQEE 1680

QY 1681 IDYDDTISVEMKKEDEDIYDEDENQSPRSFQKTRHYFIAAVERLWDYGMSSSPHVLNR 1740
DB 1681 IDYDDTISVEMKKEDEDIYDEDENQSPRSFQKTRHYFIAAVERLWDYGMSSSPHVLNR 1740
QY 1741 AQSGSVPEQFKKVVFOEFTDGSFTQPLRYGELNHLGLGPIYRAEVEDNIMVTFRCASR 1800
DB 1741 AQSGSVPEQFKKVVFOEFTDGSFTQPLRYGELNHLGLGPIYRAEVEDNIMVTFRCASR 1800
QY 1801 PYSFYSSLISYEBEQQQAEPKRNFKVKNETKTYFWKVQHMAPTKDEEDCKAWAYFSDV 1860
DB 1801 PYSFYSSLISYEBEQQQAEPKRNFKVKNETKTYFWKVQHMAPTKDEEDCKAWAYFSDV 1860
QY 1861 DLEKDVHSLIGLLVCHNTNLNPAHGRQVTVQEFALFTIIFDETYSWFTENMENCRA 1920
DB 1861 DLEKDVHSLIGLLVCHNTNLNPAHGRQVTVQEFALFTIIFDETYSWFTENMENCRA 1920
QY 1921 PCNIQMEDPTFKENYRFHAINGYIMDTLPGLVMAQDQRIRWYLLSMGSNENIHSIFSGH 1980
DB 1921 PCNIQMEDPTFKENYRFHAINGYIMDTLPGLVMAQDQRIRWYLLSMGSNENIHSIFSGH 1980
QY 1981 VFTVRKEEYKMALYNLYPGVFETVEMLPKSKAGIWRVECLIGEHHLHAGMSTIPLVYSNKC 2040
DB 1981 VFTVRKEEYKMALYNLYPGVFETVEMLPKSKAGIWRVECLIGEHHLHAGMSTIPLVYSNKC 2040
QY 2041 QTPLGMAHGHIRDFQITASGOYGOWAPKLARLHYSGSINAWSTKBPFSWKVDLLAPMII 2100
DB 2041 QTPLGMAHGHIRDFQITASGOYGOWAPKLARLHYSGSINAWSTKBPFSWKVDLLAPMII 2100
QY 2101 HGKTCGARQKTSLSVISQFIINYSLDGKKWQTYRGNSTGTLMVFFGNVDSSGIKENIEN 2160
DB 2101 HGKTCGARQKTSLSVISQFIINYSLDGKKWQTYRGNSTGTLMVFFGNVDSSGIKENIEN 2160
QY 2161 PPIIARYIELHPTHYSIRSTLRMELMGCDLNSCSPGLMGESKAISSAQITASSYFTNMEA 2220
DB 2161 PPIIARYIELHPTHYSIRSTLRMELMGCDLNSCSPGLMGESKAISSAQITASSYFTNMEA 2220
QY 2221 TWSPSKARLHLOGRSNAPVQNNPKWLOVDFOKTMKVTGVTTCQVKSLTSMVYKVEL 2280
DB 2221 TWSPSKARLHLOGRSNAPVQNNPKWLOVDFOKTMKVTGVTTCQVKSLTSMVYKVEL 2280
QY 2281 ISSSQDGHQWTLFPQNGKVKVFGNQDSTPPVNSLDPPLTRYLRIHPQSWVHQIALRM 2340
DB 2281 ISSSQDGHQWTLFPQNGKVKVFGNQDSTPPVNSLDPPLTRYLRIHPQSWVHQIALRM 2340
QY 2341 EVLGCEAODLY 2351
DB 2341 EVLGCEAODLY 2351
RESULT 13
AAW11362
ID AAW11362 standard; protein; 2351 AA.
XX
AC AAW11362;
XX
DT 18-NOV-1997 (first entry)
XX
DE Active Factor VIII:C analogue L277X.
XX
KW Factor VIII:C; analogue; glycoprotein; blood coagulation cascade;
KW fibrinogen; fibrin clot; haemostasis; haemophilia A; bleeding diathesis;
KW plasma protease; thrombin; immunogen; antibody; haemophiliac; therapy;
XX proteolytic cleavage.
OS Homo sapiens.
OS Synthetic.
XX
FH Key
FT Peptide
FT Protein
FT Region
/note= "signal peptide"
/note= "mature Factor VIII:C"
20..2351
20..1667

FT Modified-site /note= "heavy chain fragment"
 FT 296
 FT /label= pte, Glu, Pro
 FT Domain 760..1667
 FT /note= "3 domain"
 FT Region 1668..2350
 FT /note= "light chain fragment"
 XX
 XX WO9703195-A1.
 PN
 XX 30-JAN-1997.
 XX
 XX 09-JUL-1996;
 XX 96WO-US011444.
 XX
 XX 11-JUL-1995;
 XX 95US-0001025P.
 XX
 XX (CHIR) CHIRON CORP.
 PA
 XX Hung DT, Cohen FB, Innis M;
 PI WPI; 1997-119050/11.
 DE
 XX
 XX Factor VIII:C analog modified adjacent to a non-activating Arg residue -
 PT used in the treatment of haemophilias, by improvement of haemostasis.
 XX
 XX Claim 14; Page; 90pp; English.
 XX
 CC AAW11330-W11472 represent active Factor VIII:C analogues of the
 CC invention. These sequences were created by mutating the wild type Factor
 CC VIII:C coding sequence (see AAW1357) using mutagenic primers. The
 CC analogues comprise a native Factor VIII:C polypeptide modified at a site
 CC adjacent to a non-activating Arg residue so that a Arg-Pro or Pro-Arg
 CC dipeptide is created. Factor VIII:C is a large glycoprotein that
 CC participates in the blood coagulation cascade that ultimately converts
 CC soluble fibrinogen to insoluble fibrin clot, effecting haemostasis. A
 CC deficiency in Factor VIII:C is responsible for haemophilia A, which is an
 CC X-chromosome-linked inherited bleeding diathesis. Factor VIII:C is
 CC activated by plasma proteases, such as thrombin. During activation the
 CC mature polypeptide is cleaved to generate heavy and light chain fragments
 CC that are further cleaved. Complexes of two or more of the analogues,
 CC nucleic acids and vectors encoding them may be used alone or in
 CC conjunction with each other, for the prevention or treatment of active
 CC Factor VIII:C deficiency in a mammal. The analogues may be used as
 CC immunogens to raise antibodies, and in the treatment of haemophilias, by
 CC improvement of haemostasis. The analogues are resistant to proteolytic
 CC cleavage and display increased plasma half-life. They may be administered
 CC at lower dosages and by different modes of administration
 XX
 XX Sequence 2351 AA;
 SQ
 Query Match 99.8%; Score 12394; DB 2; Length 2351;
 Best Local Similarity 99.8%; Pred. No. 0;
 Matches 2347; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
 QY 1 MQIELSTCFECLLPFCPSATRYVLGAVELSWDMQSDGLGELPVDASPPRPVPSFPFN 60
 DB 1 MQIELSTCFECLLPFCPSATRYVLGAVELSWDMQSDGLGELPVDASPPRPVPSFPFN 60
 QY 61 TSVVYKTLFVEFTDHLFNIAPRPPWVGLLGPPTQAEVYDVTWITLKNMASHPVSLAV 120
 DB 61 TSVVYKTLFVEFTDHLFNIAPRPPWVGLLGPPTQAEVYDVTWITLKNMASHPVSLAV 120
 QY 121 GVSYWKASEGABYDQTSQREKEDKVPFGGSHTYVMVYKENGPMASDPICLTYSYLSH 180
 DB 121 GVSYWKASEGABYDQTSQREKEDKVPFGGSHTYVMVYKENGPMASDPICLTYSYLSH 180
 QY 181 VDLVKDLSGLIGALLVCREGSLAKEKTQTLHKFTILLFAVDEGKSWSEKNSLMQDRD 240
 DB 181 VDLVKDLSGLIGALLVCREGSLAKEKTQTLHKFTILLFAVDEGKSWSEKNSLMQDRD 240
 QY 241 AASARAWPKMHTVNGVYVNSLPGLIGCHRKSVYWHVIGMGTTPPEVHSIFLEGHTFLVRNH 300
 DB 241 AASARAWPKMHTVNGVYVNSLPGLIGCHRKSVYWHVIGMGTTPPEVHSIFLEGHTFLVRNH 300

QY 301 PQASLEISPIITFLTAQTLLMDLGOPLLSCHISSHOHDGMEAVVYVDSQPERPQLMKNE 360
 DB 301 PQASLEISPIITFLTAQTLLMDLGOPLLSCHISSHOHDGMEAVVYVDSQPERPQLMKNE 360
 QY 361 EADYDDDLTDSMDVVRFDNDSPSFTQIRSAVKGPKTWVHYIAAEEEDWDYAPLVLA 420
 DB 361 EADYDDDLTDSMDVVRFDNDSPSFTQIRSAVKGPKTWVHYIAAEEEDWDYAPLVLA 420
 QY 421 PDRSYKSOYLNNNGPQIRGRKYKVRFWAYTDETFKTREAIQHESGILGPLLYGEGVDTL 480
 DB 421 PDRSYKSOYLNNNGPQIRGRKYKVRFWAYTDETFKTREAIQHESGILGPLLYGEGVDTL 480
 QY 481 LIIFKNQASRPVNIYPHGITDVRPLYSRRLPKGVKHLKDFPILPCEIPIKYKWTVTVEGDP 540
 DB 481 LIIFKNQASRPVNIYPHGITDVRPLYSRRLPKGVKHLKDFPILPCEIPIKYKWTVTVEGDP 540
 QY 541 TKSDPRCLTRYYSFVNMRDLASGLIGPLLCYKESVDQKGNQIMSDKRNVLFSVFDE 600
 DB 541 TKSDPRCLTRYYSFVNMRDLASGLIGPLLCYKESVDQKGNQIMSDKRNVLFSVFDE 600
 QY 601 NRSWYLTENIOREFLPNPAGVQLEDEPEFOASNTMHSINGVYFDSLSQVCLHEVAWYVILS 660
 DB 601 NRSWYLTENIOREFLPNPAGVQLEDEPEFOASNTMHSINGVYFDSLSQVCLHEVAWYVILS 660
 QY 661 IGAQTDFLSVFPFGCYTFKHKMWYEDTLTFPFGSETVFMSENPGILWILGCHNSDFNRG 720
 DB 661 IGAQTDFLSVFPFGCYTFKHKMWYEDTLTFPFGSETVFMSENPGILWILGCHNSDFNRG 720
 QY 721 MTALLKVSQCKXNGDYEDSYEDISAYLLSKNNAIEPRSPQNSRHPSTQKQFNATTI 780
 DB 721 MTALLKVSQCKXNGDYEDSYEDISAYLLSKNNAIEPRSPQNSRHPSTQKQFNATTI 780
 QY 781 PENDIEKTDPMFAKRTMPKIQNVSSSLLMLLRQSPTPHGLSLSDLOEAKYETFSDDPS 840
 DB 781 PENDIEKTDPMFAKRTMPKIQNVSSSLLMLLRQSPTPHGLSLSDLOEAKYETFSDDPS 840
 QY 841 PGADSNNSLSEMTFRPQLHSGEMVFTPRSGQLRLNEKLGTTAATLKKLDPKVST 900
 DB 841 PGADSNNSLSEMTFRPQLHSGEMVFTPRSGQLRLNEKLGTTAATLKKLDPKVST 900
 QY 901 SNLITSTPSNLAAAGTDNTSSLPSPVHYDVSQDPTTLFGKSSPLTESGGPLSLSEE 960
 DB 901 SNLITSTPSNLAAAGTDNTSSLPSPVHYDVSQDPTTLFGKSSPLTESGGPLSLSEE 960
 QY 961 NNDKSLLESGLMNSQESSWGKNVSTESGRLPFGKRAHGPALLTKDNALFKVSISSLKTN 1020
 DB 961 NNDKSLLESGLMNSQESSWGKNVSTESGRLPFGKRAHGPALLTKDNALFKVSISSLKTN 1020
 QY 1021 KTSNNSATNRKTHIDGPSLLIENSPSVWQNILESDETFKVTPLIHDRMLMDKNATALRL 1080
 DB 1021 KTSNNSATNRKTHIDGPSLLIENSPSVWQNILESDETFKVTPLIHDRMLMDKNATALRL 1080
 QY 1081 NEMSNKTTSSKNMVMQOKGEPIDPAQNPDMAFFKMLFLPESARWITQTHGKNSLNSG 1140
 DB 1081 NEMSNKTTSSKNMVMQOKGEPIDPAQNPDMAFFKMLFLPESARWITQTHGKNSLNSG 1140
 QY 1141 QGPSPKQIVSLGPKSVGEGQNFLEKKNVVGKGEFTKVGLEKEMVFPSSRNLFNLN 1200
 DB 1141 QGPSPKQIVSLGPKSVGEGQNFLEKKNVVGKGEFTKVGLEKEMVFPSSRNLFNLN 1200
 QY 1201 LHENNTNQEKKIQEEIEKKEKTLIQENVLPQIHVTGTGKNFMKNLFLLSQNVGEGSYD 1260
 DB 1201 LHENNTNQEKKIQEEIEKKEKTLIQENVLPQIHVTGTGKNFMKNLFLLSQNVGEGSYD 1260
 QY 1261 GAYAPVLQDPFSLNDSINRTKKTAAHFKKGEENLGLGNQTKQIVKACTTRISGNT 1320
 DB 1261 GAYAPVLQDPFSLNDSINRTKKTAAHFKKGEENLGLGNQTKQIVKACTTRISGNT 1320
 QY 1321 SQQNFVTRSKRALKQFRLPJEETELEKRIIVDDTSTQWSKNMKHLTPSTLTQIDYNEKE 1380
 DB 1321 SQQNFVTRSKRALKQFRLPJEETELEKRIIVDDTSTQWSKNMKHLTPSTLTQIDYNEKE 1380

```
QY 1381 KGAITQSPDCLTRSHSIPQANRSPPIAKVSSFPSPRIPIYLTRVLFQDNGSHLPAASY 1440
Db 1381 KGAITQSPDCLTRSHSIPQANRSPPIAKVSSFPSPRIPIYLTRVLFQDNGSHLPAASY 1440
QY 1441 RKDSGVQSSSHPLQAKKNN:SLAILTLBMTCGQREVGSLGTSATNSVYKKVENTVLP 1500
Db 1441 RKDSGVQSSSHPLQAKKNN:SLAILTLBMTCGQREVGSLGTSATNSVYKKVENTVLP 1500
QY 1501 KPDLPTSGKVELLPKVHIYQKDLFPFET:SNQSPGHLD:VEGSLLOGTEGAIKWEANRP 1560
Db 1501 KPDLPTSGKVELLPKVHIYQKDLFPFET:SNQSPGHLD:VEGSLLOGTEGAIKWEANRP 1560
QY 1561 GKVPFLRVATSSAKTPSKLLDPLAWNDHYGTQIPKEWKSKQSKSEKTAFFKDKDTILSL 1620
Db 1561 GKVPFLRVATSSAKTPSKLLDPLAWNDHYGTQIPKEWKSKQSKSEKTAFFKDKDTILSL 1620
QY 1621 NACESNHAIAANEGONKBEIEVTWAKQGRTERLCSQNPVLRKHOREITRITLOSQDQE 1680
Db 1621 NACESNHAIAANEGONKBEIEVTWAKQGRTERLCSQNPVLRKHOREITRITLOSQDQE 1680
QY 1681 IDYDDTISVEMKKEDEFDIYDEBENQSPRSPQKTRHYFIAAVERLWDYGMSSSPHVLNR 1740
Db 1681 IDYDDTISVEMKKEDEFDIYDEBENQSPRSPQKTRHYFIAAVERLWDYGMSSSPHVLNR 1740
QY 1741 AQSGSVQPKKVVQSFDTGSGFTQPIYRGELNEHLGLGPIYRAEVEDNIMVTFRNQASR 1800
Db 1741 AQSGSVQPKKVVQSFDTGSGFTQPIYRGELNEHLGLGPIYRAEVEDNIMVTFRNQASR 1800
QY 1801 PYSFYSLISYEDDOQAGPRKFNFKNETKTYFMKVOHNAPTDREDCKAWAYFSV 1860
Db 1801 PYSFYSLISYEDDOQAGPRKFNFKNETKTYFMKVOHNAPTDREDCKAWAYFSV 1860
QY 1861 DLEKDVHSLGILGLVCHTNTLNPAGROVTVQEPALFFTFIDETKSWTFTEMNERNCA 1920
Db 1861 DLEKDVHSLGILGLVCHTNTLNPAGROVTVQEPALFFTFIDETKSWTFTEMNERNCA 1920
QY 1921 PCNIQMEDTFKFNYPFAHNGVIMDTLPGLVNAQDQIRWYLLSGNSNETHSIHFSGH 1980
Db 1921 PCNIQMEDTFKFNYPFAHNGVIMDTLPGLVNAQDQIRWYLLSGNSNETHSIHFSGH 1980
QY 1981 VFTVRKKEVKALYNLYPGVRETVMLPSKAGIWEVECLIGHLHAGMST:FLVYSNKC 2040
Db 1981 VFTVRKKEVKALYNLYPGVRETVMLPSKAGIWEVECLIGHLHAGMST:FLVYSNKC 2040
QY 2041 QTPLGWASGHIRDFOITASQYQGWAPKLARLHYSGSINAWSTKBPFSWKVDLLAPMII 2100
Db 2041 QTPLGWASGHIRDFOITASQYQGWAPKLARLHYSGSINAWSTKBPFSWKVDLLAPMII 2100
QY 2101 HGIKTQCARQKPSLLYSQPIIMYSLDGKKWQYRGNSGTGLMVFFGNVDSSGIKENIFN 2160
Db 2101 HGIKTQCARQKPSLLYSQPIIMYSLDGKKWQYRGNSGTGLMVFFGNVDSSGIKENIFN 2160
QY 2161 PPIIARYIRLPHTHYSIRSTLRMLMGCDLNSCMPLGMSXALSDAQITASSYFTNFA 2220
Db 2161 PPIIARYIRLPHTHYSIRSTLRMLMGCDLNSCMPLGMSXALSDAQITASSYFTNFA 2220
QY 2221 TWSPSXARLHLOGRSNAWRPQVNNPKXELQVDFQKTKMVTGVTQGVKSLLTSMYVKEFL 2280
Db 2221 TWSPSXARLHLOGRSNAWRPQVNNPKXELQVDFQKTKMVTGVTQGVKSLLTSMYVKEFL 2280
QY 2281 ISSSQDGHQWTLFFQNGKVKVFGNODSFTPVVNSLDPPLLTRYLRILHIFQSWVHQIALRM 2340
Db 2281 ISSSQDGHQWTLFFQNGKVKVFGNODSFTPVVNSLDPPLLTRYLRILHIFQSWVHQIALRM 2340
QY 2341 EVLGCEAQDLY 2351
Db 2341 EVLGCEAQDLY 2351
```

RESULT 14

AAW11461

ID AAW11461 standard; prote:n; 2351 AA.

XX

```
AC AAW11461;
XX 20-NOV-1997 (first entry)
DE Active Factor VIII:C analogue V1717X.
XX
KW Factor VIII:C; analogue; glycoprotein; blood coagulation cascade;
KW fibrinogen; fibrin clot; haemostasis; haemophilia A; bleeding diathesis;
KW plasma protease; thrombin; immunogen; antibody; haemophilic; therapy;
KW proteolytic cleavage.
XX
OS Homo sapiens.
XX Synthetic.
EH Key
FT Peptide
FT Protein
FT Region
FT Domain
FT Region
FT Modified-site
FT
XX W09703195-AA1.
PN 30-JAN-1997.
XX
XX 09-JUL-1996; 96WO-US011444.
XX
XX 11-JUL-1995; 95US-0001025P.
XX (CHIR ) CHIRON CORP.
XX Hung DT, Cohen PE, Innis M;
XX WPI; 1997-119050/11.
XX
XX Factor VIII:C analog modified adjacent to a non-activating Arg residue -
XX used in the treatment of haemophilias, by improvement of haemostasis.
XX
XX Claim 36; Page; 90pp; English.
XX
XX AAW11330-W11472 represent active Factor VIII:C analogues of the
XX invention. These sequences were created by mutating the wild type Factor
XX VIII:C coding sequence (see AAT51357) using mutagenic primers. The
XX analogues comprise a native Factor VIII:C polypeptide modified at a site
XX adjacent to a non-activating Arg residue so that a Arg-Pro or Pro-Arg
XX dipeptide is created. Factor VIII:C is a large glycoprotein that
XX participates in the blood coagulation cascade that ultimately converts
XX soluble fibrinogen to insoluble fibrin clot, effecting haemostasis. A
XX deficiency in Factor VIII:C is responsible for haemophilia A, which is an
XX X-chromosome-linked inherited bleeding diathesis. Factor VIII:C is
XX activated by plasma proteases, such as thrombin. During activation the
XX mature polypeptide is cleaved to generate heavy and light chain fragments
XX that are further cleaved. Complexes of two or more of the analogues,
XX nucleic acids and vectors encoding them may be used alone or in
XX conjunction with each other, for the prevention or treatment of active
XX Factor VIII:C deficiency in a mammal. The analogues may be used as
XX immunogens to raise antibodies, and in the treatment of haemophilias, by
XX improvement of haemostasis. The analogues are resistant to proteolytic
XX cleavage and display increased plasma half-life. They may be administered
XX at lower dosages and by different modes of administration
XX
XX Sequence 2351 AA;
```

Query Watch

Best Local Similarity 99.8%; Score 12394; DB 2; Length 2351;

Matches 2347; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

1 MOELSTCFCLLRFCFSATRYVYLGAVELSDYMOQSDLGELVDARPPBPVPSKPPFN 60
 Db 1 MOELSTCFCLLRFCFSATRYVYLGAVELSDYMOQSDLGELVDARPPBPVPSKPPFN 60
 QY 61 TSVVYKTLFVFTDHLFNIAKRPMPMGLGPTIOAEVYDVVITLKNASHFVLSHAV 120
 Db 61 TSVVYKTLFVFTDHLFNIAKRPMPMGLGPTIOAEVYDVVITLKNASHFVLSHAV 120
 QY 121 GYSYKASGAEYDDOTSOREKEDXVFFGGSHTYVMQVLKENGPNASDPLCLITYSLSH 180
 Db 121 GYSYKASGAEYDDOTSOREKEDXVFFGGSHTYVMQVLKENGPNASDPLCLITYSLSH 180
 QY 181 VDLVKDLSGLIGALLVCEGSLAKEKTOILHFIILLFAVDFGKSWHSETKNSLQDRED 240
 Db 181 VDLVKDLSGLIGALLVCEGSLAKEKTOILHFIILLFAVDFGKSWHSETKNSLQDRED 240
 QY 241 AASARAWPKMHTVNGYVNSRLPGLIGHRKSVYVHVI GMGTTPEVHSIFLEGHTFLVRNH 300
 Db 241 AASARAWPKMHTVNGYVNSRLPGLIGHRKSVYVHVI GMGTTPEVHSIFLEGHTFLVRNH 300
 QY 301 ROASLEISPIITFLTAOTLLMDLQGLLSCHISSEHQHDGMEAYVYKVDSCPEEPOLIKKNE 360
 Db 301 ROASLEISPIITFLTAOTLLMDLQGLLSCHISSEHQHDGMEAYVYKVDSCPEEPOLIKKNE 360
 QY 361 EABDYDDDLTDSMDVVRFDONSPSFIQIRSVAKKPKTWVHYIAAEBEDWDYAPLVLA 420
 Db 361 EABDYDDDLTDSMDVVRFDONSPSFIQIRSVAKKPKTWVHYIAAEBEDWDYAPLVLA 420
 QY 421 PDRSYKSYLNNGPORIGRYKKVRFMAVTDTPKTREAIQHESGILGPLLYGEVGDTL 480
 Db 421 PDRSYKSYLNNGPORIGRYKKVRFMAVTDTPKTREAIQHESGILGPLLYGEVGDTL 480
 QY 481 LIIFKMOASRPYNIYPHGITDVRPLYSRRLPKGVKHLKOPPIPLPGTIFKYKXTVTVEDGP 540
 Db 481 LIIFKMOASRPYNIYPHGITDVRPLYSRRLPKGVKHLKOPPIPLPGTIFKYKXTVTVEDGP 540
 QY 541 TKSDPRCLTRYSSFYNNMERDLASGLIGPLLI CYKESVDQGNQIMSKDNVILFSVDFE 600
 Db 541 TKSDPRCLTRYSSFYNNMERDLASGLIGPLLI CYKESVDQGNQIMSKDNVILFSVDFE 600
 QY 601 NRSWYLITENIQLRPLNPAGVQLEDPEFOASINIMESINGYVDFSLOQLSVCLHEVAYWYILS 660
 Db 601 NRSWYLITENIQLRPLNPAGVQLEDPEFOASINIMESINGYVDFSLOQLSVCLHEVAYWYILS 660
 QY 661 ICAQTDPLSVFFSGYTFKHQWYEDTLTLPFPGSETVFMENPGLWILGCHNSDFRNRG 720
 Db 661 ICAQTDPLSVFFSGYTFKHQWYEDTLTLPFPGSETVFMENPGLWILGCHNSDFRNRG 720
 QY 721 MTALLKVSCKDNTGDIYEDSYEDI SAYLLSKNNA-EPKSPQNSRHPSTROKQFNATTI 780
 Db 721 MTALLKVSCKDNTGDIYEDSYEDI SAYLLSKNNA-EPKSPQNSRHPSTROKQFNATTI 780
 QY 781 PENDIEKTDPFWFAHRTMPKIQNVSSDGLMLRQSTPHGLSLDLOEAKYETFSDDPS 840
 Db 781 PENDIEKTDPFWFAHRTMPKIQNVSSDGLMLRQSTPHGLSLDLOEAKYETFSDDPS 840
 QY 841 PGAI DSNNSLSEMTFRPQLHSGDMVFTPESGQLRLNEKLGTTAATLKKLDPKVSST 900
 Db 841 PGAI DSNNSLSEMTFRPQLHSGDMVFTPESGQLRLNEKLGTTAATLKKLDPKVSST 900
 QY 901 SNNLISTIPSDNLAAGTDNTSSIGPSPMPVYQSQDITLFGKKSGLTBSGGPLSLSEE 960
 Db 901 SNNLISTIPSDNLAAGTDNTSSIGPSPMPVYQSQDITLFGKKSGLTBSGGPLSLSEE 960
 QY 961 NNDKSLLESGLMNSQSSWGKNVSS-ESGRLFKGKAHGPALLTKONALFKVSI LKTN 1020
 Db 961 NNDKSLLESGLMNSQSSWGKNVSS-ESGRLFKGKAHGPALLTKONALFKVSI LKTN 1020
 QY 1021 KTSNNATNRKTHIDGPSLILIIENSPSVWQNLBSDTFEKKVTPLIIDRLMLMDKNATALL 1080
 Db 1021 KTSNNATNRKTHIDGPSLILIIENSPSVWQNLBSDTFEKKVTPLIIDRLMLMDKNATALL 1080

1081 NEMSNKTTSSKNMEXVQOKKEGPIPPDAONPDMSPFPMFLPESAEWIORTEKKSLSNG 1140
 Db 1081 NEMSNKTTSSKNMEXVQOKKEGPIPPDAONPDMSPFPMFLPESAEWIORTEKKSLSNG 1140
 QY 1141 QGSPKQLVSLGPEKSVEGQNFSLSKNKVVVGKBEFTKDVGLKEMVFPSSRLJFTNLND 1200
 Db 1141 QGSPKQLVSLGPEKSVEGQNFSLSKNKVVVGKBEFTKDVGLKEMVFPSSRLJFTNLND 1200
 QY 1201 LHENNTHOEKIKOBEIEKKEETLIQENVVLPOIHVITGTFKNPMKULFLLSTQNVEGSD 1260
 Db 1201 LHENNTHOEKIKOBEIEKKEETLIQENVVLPOIHVITGTFKNPMKULFLLSTQNVEGSD 1260
 QY 1261 GAYAPVLODFRSLNDSTNETKHTAHFSKKGSEENLEGLGNOTKQIVKIVACTTRISPN 1320
 Db 1261 GAYAPVLODFRSLNDSTNETKHTAHFSKKGSEENLEGLGNOTKQIVKIVACTTRISPN 1320
 QY 1321 SQONFVQSSKALQKFRLEETELEKRIIVDDTSTOKSKNMKHLTPSTLTQIDYNEKE 1380
 Db 1321 SQONFVQSSKALQKFRLEETELEKRIIVDDTSTOKSKNMKHLTPSTLTQIDYNEKE 1380
 QY 1381 KGAITOSPUSDCLTRSHSIPQANRSPPIAKYSSPSPRIPIYLVTRVLPQDNSSHUPAAS 1440
 Db 1381 KGAITOSPUSDCLTRSHSIPQANRSPPIAKYSSPSPRIPIYLVTRVLPQDNSSHUPAAS 1440
 QY 1441 RKDQSGVQSSSHFLQGAKKNNLSLAILTLEMTGDQREVSGLSGTSATNSVTVYKVENTVLP 1500
 Db 1441 RKDQSGVQSSSHFLQGAKKNNLSLAILTLEMTGDQREVSGLSGTSATNSVTVYKVENTVLP 1500
 QY 1501 KPDLPKTSQKVBLLPKVHIYQKDLPTTSTNSGSPGHLDLVEGSLLOGTGGAIKWNEANRP 1560
 Db 1501 KPDLPKTSQKVBLLPKVHIYQKDLPTTSTNSGSPGHLDLVEGSLLOGTGGAIKWNEANRP 1560
 QY 1561 GKVPFLVATATESAKTPSKLLDPLANDNHYGTQIKPEEMKSOEKSEKPAFKKOTILSL 1620
 Db 1561 GKVPFLVATATESAKTPSKLLDPLANDNHYGTQIKPEEMKSOEKSEKPAFKKOTILSL 1620
 QY 1621 NACESNHAAIANEGONKPEIEVWAKQGTRELCSQNPVLKRRHOREITRITLQSDQEE 1680
 Db 1621 NACESNHAAIANEGONKPEIEVWAKQGTRELCSQNPVLKRRHOREITRITLQSDQEE 1680
 QY 1681 IYDDTISVEMKKEPDIYDEBENOSPRFOKTRHYFTAAVERLMDYQMSSPHVLRRN 1740
 Db 1681 IYDDTISVEMKKEPDIYDEBENOSPRFOKTRHYFTAAVERLMDYQMSSPHVLRRN 1740
 QY 1741 AOSGSPQPKVVFQFTDGSFTQPLRYGELNEHGLLGPYIRAEVEDNIMVTFRQASR 1800
 Db 1741 AOSGSPQPKVVFQFTDGSFTQPLRYGELNEHGLLGPYIRAEVEDNIMVTFRQASR 1800
 QY 1801 PYSFYSLLISYBEDQRQGAEPKPNFVKPNETKTYFWKVQHMAPTKDEDFCKAWAYFSDV 1860
 Db 1801 PYSFYSLLISYBEDQRQGAEPKPNFVKPNETKTYFWKVQHMAPTKDEDFCKAWAYFSDV 1860
 QY 1861 DLEKDVHSGLIQPLLVCHTNTLNPAGROVTVQEPALPTIIPDEKSWYFTENMERCSA 1920
 Db 1861 DLEKDVHSGLIQPLLVCHTNTLNPAGROVTVQEPALPTIIPDEKSWYFTENMERCSA 1920
 QY 1921 PCNIQWEDTFFKBNRFAHNGYIMDTLPLVMAODORIRWYLLISMGSENTHSIFPSGH 1980
 Db 1921 PCNIQWEDTFFKBNRFAHNGYIMDTLPLVMAODORIRWYLLISMGSENTHSIFPSGH 1980
 QY 1981 VFTVRKKEYSKALXNLYPGVFTVEMLPFSKAGIWRVECLIGEHLHAGMSTLFLVYSNKC 2040
 Db 1981 VFTVRKKEYSKALXNLYPGVFTVEMLPFSKAGIWRVECLIGEHLHAGMSTLFLVYSNKC 2040
 QY 2041 QTPFLMGASHIIRDFQITASGOYQOWAPKLARLHYSGSINAWSTKEPFSWKVYDLAPMII 2100
 Db 2041 QTPFLMGASHIIRDFQITASGOYQOWAPKLARLHYSGSINAWSTKEPFSWKVYDLAPMII 2100
 QY 2101 HGKIQGARQKSSLYISQFIIMYSIDGKKWOTYRGNSTGTLMVFFGNVSSGIKHNIEN 2160
 Db 2101 HGKIQGARQKSSLYISQFIIMYSIDGKKWOTYRGNSTGTLMVFFGNVSSGIKHNIEN 2160
 QY 2161 PPIIARYIRLHPHYTHYSIRSTLEMLMGCDLNSCSMPLGMESKAISDAQITASSYFTNMFA 2220

Db 2161 PPIIARYIRLPHYSIRSLTLMELMGCDLNSCMFLGWESKAISDAQITASSYFTWMA 2220
QY 2221 TWSPSKARLHLOGRSNAWRPQVNNPKEWLQVDFQKTMKVTGVTQGVKSLTSMYVKEFL 2280
Db 2221 TWSPSKARLHLOGRSNAWRPQVNNPKEWLQVDFQKTMKVTGVTQGVKSLTSMYVKEFL 2280
QY 2281 TSSSQDGHQWTLFFQNGKVKVFGQNGDSFTPVVNSLDPPLTRYLRHQPQSWHQIALRM 2340
Db 2281 TSSSQDGHQWTLFFQNGKVKVFGQNGDSFTPVVNSLDPPLTRYLRHQPQSWHQIALRM 2340
QY 2341 EVLGCEAQDLY 2351
Db 2341 EVLGCEAQDLY 2351

RESULT 15
AAW11387
ID AAW11387 standard; protein; 2351 AA.
XX
AC AAW11387;
DT 18-NOV-1997 (first entry)
DE Active Factor VIII:C analogue N357X.
XX
KW Factor VIII:C; analogue; glycoprotein; blood coagulation cascade;
KW fibrinogen; fibrin clot; haemostasis; haemophilia A; bleeding diathesis;
KW plasma protease; thrombin; immunogen; antibody; haemophilic;
KW proteolytic cleavage.
XX
OS Homo sapiens.
OS Synthetic.
XX
FH Key Location/Qualifiers
FT Peptide 1..19
FT Protein /note= "signal peptide"
FT Protein 20..2351
FT Region /note= "mature Factor VIII:C"
FT Region 20..1667
FT Modified-site 3% /label= "heavy chain fragment"
FT Domain 760..1667
FT Domain /note= "B domain"
FT Region 1668..2350
FT Region /note= "light chain fragment"
XX
FN W09703195-A1.
XX
XX 30-JAN-1997.
XX
XX 09-JUL-1996; 96WO-US011444.
XX
XX 11-JUL-1995; 95US-0001025P.
XX
XX (CHIR) CHIRON CORP.
XX
XX Hung DT, Cohen FE, Innis M;
XX
XX WPI; 1997-119050/11.
XX
XX Factor VIII:C analog modified adjacent to a non-activating Arg residue -
XX used in the treatment of haemophilias, by improvement of haemostasis.
XX
XX Claim 20; Page; 90pp; English.
XX
XX AAW11330-W11472 represent active Factor VIII:C analogues of the
XX invention. These sequences were created by mutating the wild type Factor
XX VIII:C coding sequence (see AAT51357) using mutagenic primers. The
XX analogues comprise a native Factor VIII:C polypeptide modified at a site
XX adjacent to a non-activating Arg residue so that a Arg-Pro or Pro-Arg
XX dipeptide is created. Factor VIII:C is a large glycoprotein that

CC participates in the blood coagulation cascade that ultimately converts
CC soluble fibrinogen to insoluble fibrin clot, effecting haemostasis. A
CC deficiency in Factor VIII:C is responsible for haemophilia A, which is an
CC X-chromosome-linked inherited bleeding diathesis. Factor VIII:C is
CC activated by plasma proteases, such as thrombin. During activation the
CC mature polypeptide is cleaved to generate heavy and light chain fragments
CC that are further cleaved. Complexes of two or more of the analogues,
CC nucleic acids and vectors encoding them may be used alone or in
CC conjunction with each other, for the prevention or treatment of active
CC Factor VIII:C deficiency in a mammal. The analogues may be used as
CC immunogens to raise antibodies, and in the treatment of haemophilias, by
CC improvement of haemostasis. The analogues are resistant to proteolytic
CC cleavage and display increased plasma half-life. They may be administered
CC at lower dosages and by different modes of administration
XX
XX Sequence 2351 AA;
SQ

Query Match 99.8%; Score 12394; DB 2; Length 2351;
Best Local Similarity 99.8%; P-red. No. 0;
Matches 2347; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 MQEILSTCFCLLPFCPSATRRYVILGAVELSWDYMQSDLGELPVDARPPVPKSPFPN 60
Db 1 MQEILSTCFCLLPFCPSATRRYVILGAVELSWDYMQSDLGELPVDARPPVPKSPFPN 60
QY 61 TSVVYKTLFVEFTDHLFNIAKPRPPWMLLGPITQAEVYDTWVTLKMAASHPYSLHAV 120
Db 61 TSVVYKTLFVEFTDHLFNIAKPRPPWMLLGPITQAEVYDTWVTLKMAASHPYSLHAV 120
QY 121 GVSVMKASGARYDDQTSQREKEDKVPFGSGHTYVWQVLKENGPMASDPLCLTYSYLH 180
Db 121 GVSVMKASGARYDDQTSQREKEDKVPFGSGHTYVWQVLKENGPMASDPLCLTYSYLH 180
QY 181 VDLVKDLNSGLICALLVCREGLAKETQTLKFTLLFAVDECKSWHSETKNSLMQDRD 240
Db 181 VDLVKDLNSGLICALLVCREGLAKETQTLKFTLLFAVDECKSWHSETKNSLMQDRD 240
QY 241 AASARAWPKMHTVNGVNVNRLPGLICCHRKSVYHVI GMGTTPEVHSIFLEGTFLVRNH 300
Db 241 AASARAWPKMHTVNGVNVNRLPGLICCHRKSVYHVI GMGTTPEVHSIFLEGTFLVRNH 300
QY 301 ROASLEISPIITELTAQTLLMDLGQFLLFCHISSHODGMEAYKVVDSPCEPQLMKNE 360
Db 301 ROASLEISPIITELTAQTLLMDLGQFLLFCHISSHODGMEAYKVVDSPCEPQLMKNE 360
QY 361 EAEDYDDDLTDSEMDVWRFDNDSPSFIQIRSVAKKPKTWVHYTAAEBEDWDYAPLVIA 420
Db 361 EAEDYDDDLTDSEMDVWRFDNDSPSFIQIRSVAKKPKTWVHYTAAEBEDWDYAPLVIA 420
QY 421 PDCRSYKCYLNGKPQIRGRKYKVRFMAYTDETFKTRAIQHESGILGPLLYGEVGDTL 480
Db 421 PDCRSYKCYLNGKPQIRGRKYKVRFMAYTDETFKTRAIQHESGILGPLLYGEVGDTL 480
QY 481 LIIFKNQASRPNTYPHGTDVRLPYSRRLPKGVKHLKDPFPLPGELPKYKWTWVEDGP 540
Db 481 LIIFKNQASRPNTYPHGTDVRLPYSRRLPKGVKHLKDPFPLPGELPKYKWTWVEDGP 540
QY 541 TKSDPRCLTRYSSFVNMRDLASGLIGPLLCYKESVDQKGNMCKRNVLISVDFE 600
Db 541 TKSDPRCLTRYSSFVNMRDLASGLIGPLLCYKESVDQKGNMCKRNVLISVDFE 600
QY 601 NRSWYLTENIQRF'LNPAGVQLEDEPQASNMHSINGVYVDSLQLSVCLHEVAYWYLS 660
Db 601 NRSWYLTENIQRF'LNPAGVQLEDEPQASNMHSINGVYVDSLQLSVCLHEVAYWYLS 660
QY 661 IGAQTDFLSVFPGTFFKHKMYVETLTLFPFSGETVFMNMENPGLWILGCNDSFRNRG 720
Db 661 IGAQTDFLSVFPGTFFKHKMYVETLTLFPFSGETVFMNMENPGLWILGCNDSFRNRG 720
QY 721 MTALLKVVSSCDKNTGVDYSDSDISAYLLSKNNAIEPRSPSONSEHPSTROKQFNATTI 780
Db 721 MTALLKVVSSCDKNTGVDYSDSDISAYLLSKNNAIEPRSPSONSEHPSTROKQFNATTI 780

781	Qy	PENDIEKTDWFAHRTWPKIQNVSSDILLMLRQSPTPHGLSLDLOEAKYTFESDDPS	840
781	Db	PENDIEKTDWFAHRTWPKIQNVSSDILLMLRQSPTPHGLSLDLOEAKYTFESDDPS	840
841	Qy	PGALDSNNSLSEMTFRPQLPQHHSGDMVFTFESGLQRLNEKLGTTAATELKKLDFKVSST	900
841	Db	PGALDSNNSLSEMTFRPQLPQHHSGDMVFTFESGLQRLNEKLGTTAATELKKLDFKVSST	900
901	Qy	SNNLISTIPSDNLAAGTDNTSSGLPPSPMPVHYDSQLDPTTLFGKXSSPLTFSGGPLSLSEE	960
901	Db	SNNLISTIPSDNLAAGTDNTSSGLPPSPMPVHYDSQLDPTTLFGKXSSPLTFSGGPLSLSEE	960
961	Qy	NNDSKLLSGLMNSQESSWGKQVSYSTESGRLFKGRAGHGALLTKONALFKVSIISLLKTN	1020
961	Db	NNDSKLLSGLMNSQESSWGKQVSYSTESGRLFKGRAGHGALLTKONALFKVSIISLLKTN	1020
1021	Qy	KTSNNSATNRKTH-DGPELLIENSPPSVWQIMLESDETFKXVTPLIJHDMMLDKNATARL	1080
1021	Db	KTSNNSATNRKTHIDGPELLIENSPPSVWQIMLESDETFKXVTPLIJHDMMLDKNATARL	1080
1081	Qy	NHMSNKTTSSKNMVMQKKEGPIPPDAQNPDMGPFKMLFZLPPSARMIQTHGKXNSLNG	1140
1081	Db	NHMSNKTTSSKNMVMQKKEGPIPPDAQNPDMGPFKMLFZLPPSARMIQTHGKXNSLNG	1140
1141	Qy	QGPSPKOLVSLGPEKSVGQNFLESEKNVWVGKEFTKDVGLKEMVPPSRNULFLTLDN	1200
1141	Db	QGPSPKOLVSLGPEKSVGQNFLESEKNVWVGKEFTKDVGLKEMVPPSRNULFLTLDN	1200
1201	Qy	LHENNTNQPKKQOEBTEKXETLIOENVLIPOTHVTGTQWKNMKNLPLLSTRONVEGSD	1260
1201	Db	LHENNTNQPKKQOEBTEKXETLIOENVLIPOTHVTGTQWKNMKNLPLLSTRONVEGSD	1260
1261	Qy	GAYAPVLQDPRSLNDSNTRTKGTAHFSKKGEENLEGLGNQKQKQIIVEKYACTTRISPN	1320
1261	Db	GAYAPVLQDPRSLNDSNTRTKGTAHFSKKGEENLEGLGNQKQKQIIVEKYACTTRISPN	1320
1321	Qy	SOQFVTQSRKALKQPLPLEBTEFEKRIIVDPTSTQWKNMKNLPPSTLTQIDYNEKE	1380
1321	Db	SOQFVTQSRKALKQPLPLEBTEFEKRIIVDPTSTQWKNMKNLPPSTLTQIDYNEKE	1380
1381	Qy	KGAIQTSLSDCUTRSHSIPOANESPLPIAKVSSPFSIRPIYLITRVLFDONSSHLPAA	1440
1381	Db	KGAIQTSLSDCUTRSHSIPOANESPLPIAKVSSPFSIRPIYLITRVLFDONSSHLPAA	1440
1441	Qy	RKXDSGVQESSHFLQAGKXNMLSLAILTLEMTGDQREVGLSGTSATNSVTYKXVENTVLP	1500
1441	Db	RKXDSGVQESSHFLQAGKXNMLSLAILTLEMTGDQREVGLSGTSATNSVTYKXVENTVLP	1500
1501	Qy	KPDLPKTSGKVELLPKVHIIYOKOLPPTETSGSPGHLDLVEGSLQCTEGAIKWNANRP	1560
1501	Db	KPDLPKTSGKVELLPKVHIIYOKOLPPTETSGSPGHLDLVEGSLQCTEGAIKWNANRP	1560
1561	Qy	GKYPFLURVATESAKTSPSKLLDPLANDNHYGTQIPKEEWSQEKSPKTAFFKKDKTILSL	1620
1561	Db	GKYPFLURVATESAKTSPSKLLDPLANDNHYGTQIPKEEWSQEKSPKTAFFKKDKTILSL	1620
1621	Qy	NACESNHAIAINEGQNKPIEYTWAKQGTBLCSQNPVPLKHQBHITRITLOSQOEB	1680
1621	Db	NACESNHAIAINEGQNKPIEYTWAKQGTBLCSQNPVPLKHQBHITRITLOSQOEB	1680
1681	Qy	IDYDDTISVEMKKEDFIYDEDNQSPRSFOKKTURHYFIAAVERLMDYGMSSSPHVLNR	1740
1681	Db	IDYDDTISVEMKKEDFIYDEDNQSPRSFOKKTURHYFIAAVERLMDYGMSSSPHVLNR	1740
1741	Qy	AQSGSVQPKKVVQBEFTDGSFTQPIYRGELNBEHLGLLGPYIRAEVEDNIMVTFRQASR	1800
1741	Db	AQSGSVQPKKVVQBEFTDGSFTQPIYRGELNBEHLGLLGPYIRAEVEDNIMVTFRQASR	1800
1801	Qy	PYSFYSSLISVEDQOCAGPRXNFVKPNETKTYFWKVOHHMAPTQDEPCKKAWAYFSDV	1860
1801	Db	PYSFYSSLISVEDQOCAGPRXNFVKPNETKTYFWKVOHHMAPTQDEPCKKAWAYFSDV	1860
1861	Qy	DLEKDVHSGG-PELLVCHTNTLMPAGHRQVTVQEFALFTIFDETCKSWYFETENNERNCRA	1920

[illegible]

Search completed: April 13, 2004, 14:04:16
Job time : 77.5 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 13, 2004, 14:01:28 ; Search time 27 seconds
(without alignments)
8375.788 Million cell updates/sec

Title: NP000123-328-355-581

Perfect score: 12415

Sequence: 1 MQIELSTCFPLCLLRFCPSA.....WVHQIALRMEVLGCEAQQDLY 2351

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR 78:*

1: pir1:*

2: pir2:*

3: pir3:*

4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	12399	99.9	2351	1 EZHU	coagulation factor
2	8920	71.0	2319	2 A47004	coagulation factor
3	8057	64.9	2133	2 T42763	coagulation factor
4	2824.5	22.8	2224	1 KFRU05	coagulation factor
5	2771.5	22.3	2183	2 T42764	coagulation factor
6	2743	22.1	2211	1 KFR005	coagulation factor
7	2190	17.6	869	2 A25945	coagulation factor
8	1326	10.7	1069	1 KUHU	ferroxidase [EC 1.11.1.1]
9	1277	10.3	1059	1 A35210	ferroxidase [EC 1.11.1.1]
10	1104	8.9	216	2 A44258	factor VIII-asso
11	663	5.3	427	2 JC4915	ags protein precu
12	657	5.3	463	1 A35479	milk fat globule m
13	650	5.2	409	2 T11743	pe47 protein - pig
14	635	5.1	401	2 S65138	Glycoprotein anti
15	635	5.1	427	2 S74211	PAS-6/7 protein pr
16	443	3.6	927	1 JQ0948	A5 antigen precu
17	424.5	3.4	218	2 A47285	milk fat globule p
18	306.5	2.5	3133	2 S52093	hemocytin - silkw
19	262	2.1	845	2 JC5256	adipocyte transcr
20	239	1.9	1072	2 A86827	hypothetical prote
21	236.5	1.8	3418	1 G03234	breast cancer tumo
22	235.5	1.8	1358	2 A02930	SIRA protein - yea
23	227	1.8	2954	2 T14156	kinesin-related pr
24	224	1.8	2166	2 G70163	hypothetical prote
25	221.5	1.8	3329	2 T30904	breast cancer tumo
26	221	1.8	719	2 S51739	transcription rep
27	219.5	1.8	1928	2 S46773	myosin heavy chain
28	219.5	1.8	3329	2 T42205	breast cancer susc
29	219	1.8	2401	2 T28676	rhopty protein -

30	218.5	1.8	3507	2 T34513	hypothetical prote
31	217.5	1.8	1271	2 D64237	hypothetical prote
32	217	1.7	1381	1 S45781	probable calcium-b
33	216.5	1.7	3328	2 T30835	breast cancer tumo
34	216	1.7	3924	2 S37431	ankyrin 2, neuro
35	213.5	1.7	2346	2 T13829	TPR homolog - frui
36	213	1.7	1420	1 A44361	amiloride-sensitiv
37	212.5	1.7	5005	2 F82884	hypothetical prote
38	212	1.7	1957	2 T38077	hypothetical coile
39	211	1.7	1283	2 T13799	neurexin IV - frui
40	209.5	1.7	1177	2 T64233	hypothetical prote
41	208.5	1.7	1628	2 E30538	hypothetical prote
42	207.5	1.7	737	2 T31349	hypothetical prote
43	207	1.7	6669	2 S55024	nebulin, skeletal
44	206.5	1.7	737	2 T15615	hypothetical prote
45	204.5	1.6	1805	2 T02712	similar to late em

ALIGNMENTS

RESULT 1

EZHU
coagulation factor VIII precursor [validated] - human
N:Alternate names: antihemophilic factor A; coagulation factor VIIIc; procoagulant compon
C:Species: Homo sapiens (man)
C:Date: 28-Aug-1985 #sequence revision 28-Aug-1985 #text change 08-Dec-2000
C:Accession: I54318; A00525; I58059; A23584; A26174; A42348; A43986; S63527; S66445; B42:
R:Gitschier, J.; Wood, W.I.
Hum. Mol. Genet. 1, 199-200, 1992
A:Title: Sequence of the exon-containing regions of the human factor VIII gene.
A:Reference number: I54318; MUID:93265012; PMID:1303178
A:Accession: I54318
A:Status: preliminary; translated from GB/EMBL/DBDJ
A:Molecule type: DNA
A:Residues: 1-1921, 'S', 1923-2351 <RES>
A:Cross-references: GB:M88648; NID:G182381; PIDN:AAA52420.1; PID:G182383
R:Wood, W.I.; Capon, D.J.; Simonsen, C.C.; Eaton, D.L.; Gitschier, J.; Keyt, B.; Seeburg
Nature 312, 330-337, 1984
A:Title: Expression of active human factor VIII from recombinant DNA clones.
A:Reference number: A00525; MUID:85061548; PMID:6438526
A:Accession: A00525
A:Molecule type: mRNA
A:Residues: 1-2351 <WOO>
A:Cross-references: EMBL:X01166; EMBL:X01179
R:Toole, J.J.; Knopf, J.L.; Wozney, J.M.; Sultzman, L.A.; Buecker, J.S.; Pittman, D.D.; J
S, D.N.; Hewick, R.M.
Nature 312, 342-347, 1984
A:Title: Molecular cloning of a cDNA encoding human antihemophilic factor.
A:Reference number: I58059; MUID:85061550; PMID:6438528
A:Accession: I58059
A:Status: nucleic acid sequence not shown; translated from GB/EMBL/DBDJ
A:Molecule type: mRNA
A:Residues: 1-74, 'V', 76-259, 'E', 1361-2351 <RE2>
A:Cross-references: GB:X01740; NID:G182802; PIDN:AAA52484.1; PID:G182803
R:Truett, M.A.; Blacher, R.; Burke, R.L.; Caput, D.; Chu, C.; Dina, D.; Hartog, K.; Kuo,
B.; Randolph, A.; Urdea, M.S.; Valenzuela, P.; Dahl, H.H.; Favalaro, J.; Hansen, J.; No
DNA 4, 333-349, 1985
A:Title: Characterization of the polypeptide composition of human factor VIII:C and the
A:Reference number: A23584; MUID:86081164; PMID:3935400
A:Accession: A23584
A:Molecule type: mRNA
A:Residues: 1-2351 <TRU>
A:Cross-references: GB:M14113; NID:G182817; PIDN:AAA52485.1; PID:G182818
R:Eaton, D.; Rodriguez, H.; Vehar, G.A.
Biochemistry 25, 505-512, 1986
A:Title: Proteolytic processing of human factor VIII. Correlation of specific cleavages
ity.
A:Reference number: A26174; MUID:86159740; PMID:3082357
A:Accession: A26174
A:Molecule type: protein
A:Residues: 20-36; 392-399, 'X', 401-402; 1668-1678; 1709-1722, 'D', 1723-1725; 1741-1755 <EAT>
R:Pittman, D.D.; Wang, J.H.; Kaufman, R.J.

Biochemistry 31, 3315-3325, 1992
A;Title: Identification and functional importance of tyrosine sulfate residues within re
A;Reference number: A42348; MUID:92207952; PMID:1554716
A;Accession: A42348
A;Molecule type: protein
A;Residues: 20-36,356-371;392-408;582-594;1668-1669,'X',1671;1672-1692;1693-1708;1709-17
A;Experimental source: Recombinant material from Chinese hamster ovary cells
A;Note: sequence extracted from NCBI backbone and corrected to correspond with the publi
E;Pay, P.J.; Smudz, T.M.
J. Biol. Chem. 264, 14005-14010, 1989
A;Title: Intersubunit fluorescence energy transfer in human factor VIII.
A;Reference number: A43986; MUID:89340500; PMID:2503509
A;Accession: A43986
A;Molecule type: protein
A;Residues: 'X',517-523;1853-1860,'X',1862-1864,'X',1866 <FAY>
R;Leyte, A.; van Schijndel, H.B.; Niehrs, C.; Ruttner, W.B.; Verbeet, M.P.; Martens, K.;
J. Biol. Chem. 266, 740-746, 1991
A;Title: Sulfation of Tyr(1680) of human blood coagulation factor VIII is essential for
A;Reference number: A56109; MUID:91093266; PMID:1898735
A;Contents: annotation; sulfation
R;Gutschier, J.; Wood, W.I.; Goralka, T.M.; Wion, K.L.; Cher, E.Y.; Eaton, D.H.; Vehar,
Nature 312, 326-330, 1984
A;Title: Characterization of the human factor VIII gene.
A;Reference number: A56196; MUID:85061547; PMID:6438525
A;Contents: annotation; introns
R;McMullen, B.A.; Fujikawa, K.; Davie, E.W.; Hedner, U.; Ezban, M.
Protein Sci. 4, 740-746, 1995
A;Title: Locations of Glutamide bonds and free cysteines in the heavy and light chains o
A;Reference number: A56216; MUID:95338127; PMID:7613471
A;Contents: annotation; disulfide bonds
A;Note: 329-Cys, 711-Cys, and 2019-Cys were shown to have free sulfhydryls
R;Kjalkke, M.; Hedner, A.; Talbo, G.; Persson, E.; Thomsen, G.; Ezban, M.
Eur. J. Biochem. 234, 773-779, 1995
A;Title: Amino acid residues 721-729 are required for full factor VIII activity.
A;Reference number: S63527; MUID:96163459; PMID:8575434
A;Accession: S63527
A;Molecule type: protein
A;Residues: 733-752;753-759 <KJA>
R;Bind, P.; Larsson, K.; Spira, J.; Sydow-Baackman, M.; Ailmstedt, A.; Gray, E.; Sandberg
Eur. J. Biochem. 232, 19-27, 1995
A;Title: Novel forms of B-domain-deleted recombinant factor VIII molecules. Construction
A;Reference number: S66445; MUID:96048024; PMID:7556150
A;Accession: S66445
A;Status: preliminary
A;Molecule type: protein
A;Residues: 1668-1685 <JIN>
C;Comment: Factor VIII is activated by factor Xa and thrombin, but prolonged exposure pr
C;Genetics:
A;Gene: GDB:F8C
A;Cross-references: GDB:119124; OMIM:306700
A;Map position: Xg28-Xc28
A;Introns: 48/2; 89/1; 130/1; 201/1; 224/1; 263/1; 337/1; 424/2; 481/3; 513/1; 584/3; 63
C;Function:
A;Description: acts as a cofactor, with calcium and phospholipid, for the factor IXa pro
A;Pathway: blood coagulation
C;Superfamily: coagulation factor VIII; disocidin I amino-terminal homology; ferroxidase
C;Keywords: acute phase; blood coagulation; duplication; glycoprotein; hemophilia A; pla
F;1-19/Domain: signal sequence #status predicted <SIG>
F;20-2351/Product: coagulation factor VIII #status experimental <MAT>
F;20-740/Product: coagulation factor VIIa heavy chain #status experimental <ACE>
F;20-356/Domain: A1 <DA1>
F;23-348/Domain: A2 <DA2>
F;392-759/Domain: A3 <DA3>
F;402-730/Domain: ferroxidase repeat homology <F02>
F;760-1667/Domain: B <DB0>
F;1668-2351/Product: coagulation factor VIIa light chain #status experimental <ACL>
F;1709-2038/Domain: A3 <DA3>
F;1716-2038/Domain: ferroxidase repeat homology <F03>
F;2039-2169/Domain: C1 <DC1>
F;2039-2169/Domain: disocidin I amino-terminal homology <DN1>
F;2192-2351/Domain: C2 <DC2>
F;2192-2345/Domain: disocidin I amino-terminal homology <DN2>
F;60,258,601,776,803,847,919,962,982,1020,1024,1074,1085,1204,1274,1278,1301,1319,1403,1

F;172-198,267-348,547-573,649-730,1851-1877,1918-1922,2040-2188/Disulfide bonds: #status
F;355-356/cleavage site: Arg-Met (coagulation factor Xa, prorein C) #status predicted
F;355,727,728,742,1683,1692/Binding site: sulfate (Tyr) (covalent) #status experimental
F;351-382/cleavage site: Arg-Ser (coagulation factor Xa, thrombin) #status experimental
F;414,426/Binding site: sulfate (Tyr) (covalent) #status predicted
F;759-760/cleavage site: Arg-Ser (coagulation factor Xa, thrombin) #status experimental
F;1667-1668/cleavage site: Arg-Glu (unidentified proteinase) #status experimental
F;1708-1709/cleavage site: Arg-Ser (coagulation factor Xa, thrombin) #status experimental
F;1740-1741/cleavage site: Arg-Ala (coagulation factor Xa, thrombin) #status experimental
F;2193-2345/Disulfide bonds: #status predicted
Query Match 99.9%; Score 12399; DB 1; Length 2351;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 2348; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
QY 1 MQEILSTCFPLCLLRFCFSATRIYILGAVELSDYWDYQSDLGELPVDARFPRVPKSPFN 60
Db 1 MQEILSTCFPLCLLRFCFSATRIYILGAVELSDYWDYQSDLGELPVDARFPRVPKSPFN 60
QY 61 TSVVYKTLFVEFTDHLFNIAKPRPPMGLLGTIOAEVYDVTYVITLKNMASHPVSLHAV 120
Db 61 TSVVYKTLFVEFTDHLFNIAKPRPPMGLLGTIOAEVYDVTYVITLKNMASHPVSLHAV 120
QY 121 GVSYWKASGAEYDDQTSQREKEDDKVFPQSGSHYVQVLKENGPMASDPLCLITYSLH 180
Db 121 GVSYWKASGAEYDDQTSQREKEDDKVFPQSGSHYVQVLKENGPMASDPLCLITYSLH 180
QY 181 VDLVKOLNSGLICALLVCREGSLAKETQTLHKFILLFAVDEGKSWHSETKSLMQDRD 240
Db 181 VDLVKOLNSGLICALLVCREGSLAKETQTLHKFILLFAVDEGKSWHSETKSLMQDRD 240
QY 241 AASARAWPKMHTVNGYVNSLPLGLICHRKSVVHVGMTTPEVHSIFLEHTEFLVRNH 300
Db 241 AASARAWPKMHTVNGYVNSLPLGLICHRKSVVHVGMTTPEVHSIFLEHTEFLVRNH 300
QY 301 ROASLEISPTITLTAQLMDLQGLFLSCHISSHQDGMAYVYKVDSCPEEPQLMKNE 360
Db 301 ROASLEISPTITLTAQLMDLQGLFLSCHISSHQDGMAYVYKVDSCPEEPQLMKNE 360
QY 361 EADYDDDDTDSMDVVRDDNSPSFIOIRSVAKKPKTKWHYIAAEEDWDYAPLVLA 420
Db 361 EADYDDDDTDSMDVVRDDNSPSFIOIRSVAKKPKTKWHYIAAEEDWDYAPLVLA 420
QY 421 PDRSYKSYLNGPQIRGRKYKVRFMAYTDTFTKTRAIQHSGLIPLLYGEVGDIL 480
Db 421 PDRSYKSYLNGPQIRGRKYKVRFMAYTDTFTKTRAIQHSGLIPLLYGEVGDIL 480
QY 481 LIIFKQASRPNIYPHGTDVRLYSRRLPKGVKHLKOPPLPGEIFKYKWTVTVEDGP 540
Db 481 LIIFKQASRPNIYPHGTDVRLYSRRLPKGVKHLKOPPLPGEIFKYKWTVTVEDGP 540
QY 541 TKSDPRCLTRYSSFVNMRDLASGLIGLILCYKESVDKNGNOIMSDKENVILSVFDE 600
Db 541 TKSDPRCLTRYSSFVNMRDLASGLIGLILCYKESVDKNGNOIMSDKENVILSVFDE 600
QY 601 NRSWLTENIQRLPNPAGVQLEDEPQASNMHSINGYVFDLSQLSVCLHEVAYWYILS 660
Db 601 NRSWLTENIQRLPNPAGVQLEDEPQASNMHSINGYVFDLSQLSVCLHEVAYWYILS 660
QY 661 IGAQTDPLSVFSGYTFKHKMYEDTLTLPPSGGTVMMSNPGLWILGCHNSDFRNG 720
Db 661 IGAQTDPLSVFSGYTFKHKMYEDTLTLPPSGGTVMMSNPGLWILGCHNSDFRNG 720
QY 721 MTALLKVVSSCDKNTGDYVSDYSYLLSKNNAIEPRSPNSRHPSTROKQNNATI 780
Db 721 MTALLKVVSSCDKNTGDYVSDYSYLLSKNNAIEPRSPNSRHPSTROKQNNATI 780
QY 781 PENDIEKTDPFARHTPMPKIQNVSSDLMLLRQSPTPHGLSLDLQAKYETSDDPS 840
Db 781 PENDIEKTDPFARHTPMPKIQNVSSDLMLLRQSPTPHGLSLDLQAKYETSDDPS 840
QY 841 PGALDNNLSLMTWTFRQLHSGDMVFTPSGLQRLNEKLTGTTAAETLKLDPKVGST 900
Db 841 PGALDNNLSLMTWTFRQLHSGDMVFTPSGLQRLNEKLTGTTAAETLKLDPKVGST 900

Db 841 PGALDSNNSLSEMTFHPQLHSGDMVFTPEBSGLQLRLNEKLGTTAAATELKKCLDFKVSST 900
QY 901 SNMLISITIPSDNLAAGTNTSSIGPPSPVHYDSOLDTTLFGKSSPLTSGGPLSLSEE 960
Db 901 SNMLISITIPSDNLAAGTNTSSIGPPSPVHYDSOLDTTLFGKSSPLTSGGPLSLSEE 960
QY 961 NNDSPKLESGLMNSQESSWGKNSSTESGRLPFGKGAHGPAALLTKONALFKVSI SLLKTN 1020
Db 961 NNDSPKLESGLMNSQESSWGKNSSTESGRLPFGKGAHGPAALLTKONALFKVSI SLLKTN 1020
QY 1021 KTSNNSTNKTHTIDGSPSLIENSPPVQWNLBSDFEFKKTPLIHDRMLMDKNAATLRL 1080
Db 1021 KTSNNSTNKTHTIDGSPSLIENSPPVQWNLBSDFEFKKTPLIHDRMLMDKNAATLRL 1080
QY 1081 NMSNKTTSNNKMEMVOQKKEGTPPPDAQNPDMSPFFKMLFLPESAKWIQETHGKNSLNSG 1140
Db 1081 NMSNKTTSNNKMEMVOQKKEGTPPPDAQNPDMSPFFKMLFLPESAKWIQETHGKNSLNSG 1140
QY 1141 QGSPKQLVSLGPEKSVGQNF-SEKNKVVVGKEFTKQVGLKEMVFPSSRNFLTNLDN 1200
Db 1141 QGSPKQLVSLGPEKSVGQNF-SEKNKVVVGKEFTKQVGLKEMVFPSSRNFLTNLDN 1200
QY 1201 LHENNTNHOEKKIQLIEKKEKTELIOENVLPQIHTVTGTGNFMKQFLFLSTRQNVGSGYD 1260
Db 1201 LHENNTNHOEKKIQLIEKKEKTELIOENVLPQIHTVTGTGNFMKQFLFLSTRQNVGSGYD 1260
QY 1261 GAYAPVLQDPRSLNDSTNRKKHTAHPKSKGEBENLEGNGNQTKQIVEKYACTRISPNT 1320
Db 1261 GAYAPVLQDPRSLNDSTNRKKHTAHPKSKGEBENLEGNGNQTKQIVEKYACTRISPNT 1320
QY 1321 SQCNFTQRSKRALQKQFLPLEETELEKRIIIVDDTSTQWSKNMKHLPSTLTQIDYNEKE 1380
Db 1321 SQCNFTQRSKRALQKQFLPLEETELEKRIIIVDDTSTQWSKNMKHLPSTLTQIDYNEKE 1380
QY 1381 KGAITQSPJSDCLTRSHSIPQANRSLPATAKVSFPSPRIPIYTRVLFDQNSHSLPAASY 1440
Db 1381 KGAITQSPJSDCLTRSHSIPQANRSLPATAKVSFPSPRIPIYTRVLFDQNSHSLPAASY 1440
QY 1441 RKKDSGVQSPSSHPLOAKQKNNLSLALTLEMTGDQREVSLGTSATNSVTYKKAVENTVLP 1500
Db 1441 RKKDSGVQSPSSHPLOAKQKNNLSLALTLEMTGDQREVSLGTSATNSVTYKKAVENTVLP 1500
QY 1501 KQDLPTSGKVELLPKHVYQKDLPTTETSNQSPGHLDLVEGSLQCTEGAIKWNENRP 1560
Db 1501 KQDLPTSGKVELLPKHVYQKDLPTTETSNQSPGHLDLVEGSLQCTEGAIKWNENRP 1560
QY 1561 GXVPPFLRVATESAKTPSKLLDPLANDHVGTOIPKEENKQESPEKTAFFKKOTILSL 1620
Db 1561 GXVPPFLRVATESAKTPSKLLDPLANDHVGTOIPKEENKQESPEKTAFFKKOTILSL 1620
QY 1621 NACESNEAJAINEGONKPEIEVYTWAKQRTERLCSQNPVPLKRGHOREITRTTLQSDQEE 1680
Db 1621 NACESNEAJAINEGONKPEIEVYTWAKQRTERLCSQNPVPLKRGHOREITRTTLQSDQEE 1680
QY 1681 IDYDDTISVEMKEDPDIDYDENQSPRQFQKTRHYPTAAVERLMDYGNSSPHVLRNR 1740
Db 1681 IDYDDTISVEMKEDPDIDYDENQSPRQFQKTRHYPTAAVERLMDYGNSSPHVLRNR 1740
QY 1741 AQSGSVQPKKVVQFQFTDGSFTQPLVYRGELNEHLLGPGYIRAEVEDNIMVTFRQASR 1800
Db 1741 AQSGSVQPKKVVQFQFTDGSFTQPLVYRGELNEHLLGPGYIRAEVEDNIMVTFRQASR 1800
QY 1801 PYSFYSSLISYBEDQQAEPKRNFKVKNETKTYFWKVQHMAPTKDEPDCKAWAYFSDV 1860
Db 1801 PYSFYSSLISYBEDQQAEPKRNFKVKNETKTYFWKVQHMAPTKDEPDCKAWAYFSDV 1860
QY 1861 DLEKDVHSLGIGPLLIVCHTINTLNPAHGRQVTVQEPALFTTIDETKSWYFTENMERENCA 1920
Db 1861 DLEKDVHSLGIGPLLIVCHTINTLNPAHGRQVTVQEPALFTTIDETKSWYFTENMERENCA 1920
QY 1921 PCNIQWEDPTFKENYREHAINGYIMDTPLGLWMAQDQRIWYLLSGNSNENHSIHFSGH 1980
Db 1921 PCNIQWEDPTFKENYREHAINGYIMDTPLGLWMAQDQRIWYLLSGNSNENHSIHFSGH 1980

QY 1981 VFTVRKKEEYKMALYNLYPGVFFTEVEMLPKAGIWAIVECLIGBELHAGMTTLFLVYSNKC 2040
Db 1981 VFTVRKKEEYKMALYNLYPGVFFTEVEMLPKAGIWAIVECLIGBELHAGMTTLFLVYSNKC 2040
QY 2041 QTEPLGMASGHIRDFQITASGOYQOWAPKLARLHYSGSINAWSTKEPPSWTKVDLLAPMII 2100
Db 2041 QTEPLGMASGHIRDFQITASGOYQOWAPKLARLHYSGSINAWSTKEPPSWTKVDLLAPMII 2100
QY 2101 HGIKTQCARQKFSLSYISQFIIMYSLDGKKWQTYRGNSTGTLMVFFGNVDSSGIKENIFN 2160
Db 2101 HGIKTQCARQKFSLSYISQFIIMYSLDGKKWQTYRGNSTGTLMVFFGNVDSSGIKENIFN 2160
QY 2161 PPIIARYIRLHPTEYISIRSTLRMELMGCNLSNCSMPLGMSKAI SDAQITASSYFTNMPA 2220
Db 2161 PPIIARYIRLHPTEYISIRSTLRMELMGCNLSNCSMPLGMSKAI SDAQITASSYFTNMPA 2220
QY 2221 TWSPSKARLHLQGRSNARPOVNNPKEMIQVDPQKTMKVTVGTTQGVKSLLTSWYKEFL 2280
Db 2221 TWSPSKARLHLQGRSNARPOVNNPKEMIQVDPQKTMKVTVGTTQGVKSLLTSWYKEFL 2280
QY 2281 ISSSQDGHQWTLFFQNGKVKVFGQNDSETPPVVNSLDPPLTRYLRIHPQSWHQAIALRM 2340
Db 2281 ISSSQDGHQWTLFFQNGKVKVFGQNDSETPPVVNSLDPPLTRYLRIHPQSWHQAIALRM 2340
QY 2341 EVLGCBAQDLY 2351
Db 2341 EVLGCBAQDLY 2351

Query Match 71.0%; Score 8820; DB 2; Length 2319;

Best Local Similarity 72.4%; Pred. No. 0;

Matches 1711; Conservative 232; Mismatches 365; Indels 54; Gaps 22;

QY 1 MQIELSTCFELCLLRPCSATRRYVILGAVELSWDYMQSD-LGELPVDARPPRPVPSPPF 59

Db 1 MQIALFACFFLSLNFPCSSAIRRYVILGAVELSWNYIQSDLLSVLHTDSRFLPRMSTSPFF 60

QY 60 NTSVVYKTLFVEFTDHLFNIAKPEPPPMWGLLGPITQAEVVDVTVITLKNWASHPSVLSHA 119

Db 61 NTSIMYKTVFYVYKQDLFNIAKPEPPPMWGLLGPITWTEVHDVTVITLKNWASHPSVLSHA 120

QY 120 VGVSYKASEAGAYDDQTSQREKDDKVPFGSGSHYVWQVLKENGPMASDPLCLITYSLS 179

Db 121 VGVSYKASEGEGYEDQTSQWEKDDKVPFGSHYVWQVLKENGPMASDPLCLITYSLS 180

QY 180 HYDLVKDNLNSGLIGALLVCRRGSLAKEKTYOTLHKPILFAVFDGKSWHSTKNSLMQDR 239

Db 181 HYDLVKDNLNSGLIGALLVCRRGSLKSKERTQMLYQVLLFAVFDGKSWHSTKNSLYTQSM 240

RESULT 2

A47004

coagulation factor VIII precursor - mouse

C:Species: Mus musculus (house mouse)

C:Date: 09-Sep-1994 #sequence_revision 09-Sep-1994 #-ext_change 18-Jun-1999

C:Accession: A47004

R:Ellder, B.; Lakich, D.; Gierschier, J.

Genomics 16, 374-379, 1993

A:Title: Sequence of the murine factor VIII cDNA.

A:Reference number: A47004; MUID:93300511; PMID:8314577

A:Accession: A47004

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-2319 <EID>

A:Cross-references: GB:105573; NID:9192456; PIDN:AAA73385.1; PID:9192457

C:Superfamily: coagulation factor VIII; discolidin I amino-terminal homology; ferroxidase

F;1-19/Domain: signal sequence #status predicted <SIG>

F;23-349/Domain: ferroxidase repeat homology <FO1>

F;402-730/Domain: ferroxidase repeat homology <FO2>

F;1686-2006/Domain: ferroxidase repeat homology <FO3>

F;2007-2156/Domain: discolidin I amino-terminal homology <DN1>

F;2160-2313/Domain: discolidin I amino-terminal homology <DN2>

T42763
 coagulation factor VIII precursor - pig
 C:Species: Sus scrofa domestica (domestic pig)
 C>Date: 11-Jan-2000 #sequence_revision 11-Jan-2000 #text_change 09-Jun-2000
 C:Accession: T42763
 R:Julliar, P.
 submitted to the EMBL Data Library, August 1996
 A:Reference number: Z22269
 A:Status: preliminary; translated from G3/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1-2133 <LOL>
 A:Cross-references: EMBL:U49517; NID:g1511633; PID:g1511634; PIDN:AAB06705.1
 C:Superfamily: coagulation factor VIII; discolidin I amino-terminal homology; ferroxidase
 C:Keywords: acute phase; blood coagulation; duplication; glycoprotein; hemophilia A; pla
 F:1-19/Domain: signal sequence #status predicted <SIS>
 F:20-2133/Product: coagulation factor VIII #status predicted <MAT>
 F:23-349/Domain: ferroxidase repeat homology <FOX1>
 F:402-730/Domain: ferroxidase repeat homology <FOX2>
 F:1498-1820/Domain: ferroxidase repeat homology <FOX3>

Query Match 64.9%; Score 8057; DB 2; Length 2133;
 Best Local Similarity 66.9%; Pred. No. 0;
 Matches 1575; Conservative 209; Mismatches 345; Indels 226; Gaps 17;

QY	1	MOELSTCFLLRRCFCGATRRYLCAVELGWDYMQSD-LGELPVDAREPRVRKSPFP	59
DB	1	MOELSTCVFLCLPLGFSAIRRYLCAVELGWDYRQSELLRELVDTFPATAPALPL	60
QY	60	NTSVYKTLFVEFDHLFNIAKPPPMGLLPTIAEVYTVITLKNMASHPVSLHA	119
DB	61	GPSVLYKTVFVEFDQLPSVARPPPMGLLPTIAEVYTVITLKNMASHPVSLHA	120
QY	120	GVGSYWKASEGAEYDDTSQRKEDDKVPPGSHYVWQVLZENGPMASDPCLITYSL	179
DB	121	GVGSYWKSEGAEEYDHTSQREKEDKVLPGSKSYVWQVLZENGPMASDPCLITYSL	180
QY	180	HYDLVKDLNSGLIGALLVCREGLAKETOTLHKPILLFAVDEGKSWHSEKNSLMQDR	239
DB	181	HYDLVKDLNSGLIGALLVCREGLTPERTONLHBFVLLFAVDEGKSWHSEKNSLMQDR	240
QY	240	DAASARAWPMHTVNGYVNRSLPGLIGCHRKSYVHVIQMGTPVHSLFEGHTFLVNR	299
DB	241	DPAPAAQPMHTVNGYVNRSLPGLIGCHRKSYVHVIQMGTPVHSLFEGHTFLVNR	300
QY	300	HQAASLEISPIFLTAQTLMDLGOFLISCHTSSSHQDGEAYVXVDSCEPEPOLIMKN	359
DB	301	HQAASLEISPIFLTAQTLMDLGOFLISCHTSSSHQDGEAYVXVDSCEPEPOLIMKN	360
QY	360	EEADYDDDLTDSMDVVRFDSDNSPFTQIRSVAKKHPTVWVYIAAEEEDWDYAPVL	419
DB	361	EE-EDYDNLVSDMDVVRFDSDNSPFTQIRSVAKKHPTVWVYIAAEEEDWDYAPVL	419
QY	420	APDHSYKSOYLNNQPGRIKRYKVRNAYTDTFTKTRTAAQHSGLGLLYGEVGT	479
DB	420	SPDSYSKSLYNSQPGRIKRYKVRNAYTDTFTKTRTAAQHSGLGLLYGEVGT	479
QY	480	LTIFKQASRPYNYPHGIDTVRPLYSRLPKYKHLKDFILBGETFKYKWTTVEDG	539
DB	480	LUUIFKKASRPYNYPHGIDTVRPLYSRLPKYKHLKDFILBGETFKYKWTTVEDG	539
QY	540	PTKSDPRCLTRYSSFVNMRDLASGLIGLLICYKESVDQKGNQIMSDKRNVLFSVFD	599
DB	540	PTKSDPRCLTRYSSINLEKDLASGLIGLLICYKESVDQKGNQIMSDKRNVLFSVFD	599
QY	600	ENRSHYLTENIRFPLNPAQVQLDEPEQASNMHSINGYVDSIQLSVCLHEVAYWYL	659
DB	600	ENQSYLAENIRFPLNPAQVQLDEPEQASNMHSINGYVDSIQLSVCLHEVAYWYL	659
QY	660	SICAOITDLVSFPFGSYTPKHKVYVEDTLTLFPSETGTFMSENGLNLGLCHNSDFNR	719
DB	660	SVGAQITDLVSFPFGSYTPKHKVYVEDTLTLFPSETGTFMSENGLNLGLCHNSDFNR	719

QY	720	GNTALLKVSSCDKNTGDYEDSYEDISAYLLSKNNAIEPRSPQNSRHHSTKQKQNMAT	779
DB	720	GNTALLKVYSCDRDIGDYDNTYEDIPGLLSGKVIIEPRSPQNSRHHSTKQKQNMAT	779
QY	780	IPENDIEKTDPMFAHRTMPKIQNVSSSDLLMLLRQSPHGLSLDLQEAKEYETSDP	839
DB	780	SPEDDVE-LDPOSGERTQALELSVPSGDMGLGQNPAPPHSGSSSDLOEANE--ADY	836
QY	840	SPGAISSNNLSSEMTHFRPQLHSGDMVFTPSG+QLRLNEKLGTTAAATLKLKLPKVS	899
DB	837	LPGARERTAPGAARLRDELHSAERVJTPEP-----EK-----ELKLSKMS	882
QY	900	TNNLIS--TISDMLAAGTNTSSIGPSPVHDSQLDITLPGKSSPETSBSGPLSL	957
DB	893	SDLLKTSPTISDLSAETERTSLGPPHPPQNFERSQLGALVGLKNSHFTGAGVPLGS	942
QY	958	SENNDSKLLSGLMKNSQBSMGKNSVSTESRLEFKGAHGPALLTKDNALPKVYSILL	1017
DB	943	TEEDH-----YSSIGENVSPVSDGIFEKRAHGPASLTKDDVLFKNISLV	989
QY	1018	KTKNTSNNATNRKTHIDGPSILLIENSPSVWONILESDETFKKV77ELIHDRMLDKNATA	1077
DB	990	KTKARVYLKTKHIDDAALLTENASA-----TFMDKNTTA	1028
QY	1078	LELNHMSNKTSSKNMENVQKKGPIPPDAQNPDMSFEKMLFLPESARWIOPTHGKNSL	1137
DB	1029	SGLNHVN-----WIKGLGKNPL	1047
QY	1138	NSGQPSPKQVSLGPEKSVGONFLSKNVVGVGKFTKDVGLKEMVFPSSRLFLTN	1197
DB	1048	SSERGSPPELLTSSGSKSVKQSGQGRIRVAVSEELSKG--KEMMLPNSLFTLN	1104
QY	1198	LDNLHNNTHQEKIKOBEIEKKTLLQENVVLPQIHTVGTGNFMKNLFLSTRQNVG	1257
DB	1105	SADVQNDTHSQCKSKREMERREKLVQEKVDLPQVYATGKTNFLRNIFHQSTSPVVG	1164
QY	1258	SYDGAAPVLQDFPSLNDSTNRTKHTAHFKSGKEENLEGKNOTKQIVEXACTTRIS	1317
DB	1155	PDGGHAPVQPSFSDNSASRAETHIAHFAIRBEAPLEAFGNRI-----	1210
QY	1318	NTSQNFVTOFSKALKQFLPLEBEETELEKRIIVDDTSTQMSKMKHLPSTLTQIDVN	1377
DB	1211	-GPGPSAPRVRKQSLQIRLPLEBIEIKPGRGVNLNSTRWS-----	1252
QY	1378	EKEGAIQSPSLSDCLTRSHSIPQANRSLPIAKVSSPSPRIYLRVLPDQNSHLLPA	1437
DB	1253	-----	1252
QY	1438	ASRKKSGVQESSHFLOGAKNNLSLAILTLEMTGDQREYVSGLSGTSATNSVYKKVENT	1497
DB	1253	-----ESSFLOGAKNNLSLFLILEVAGGKGKISALGKSAAGPLASGKLEKA	1301
QY	1498	VLPKPLPKTSGKVLELKVHIYOKDLFPPTTSNGSPGHLDLVEGSLLGQTEGAIKWNEA	1557
DB	1302	VLSAGLSEASGKGFPLPKVVRHEDLLPQKTSNVSCAHGDLGQBIFLQKTRGPNLVNKV	1361
QY	1558	NPQKVPPLRVATSSAKTPSKLDPLAWDNYHTQIPKEWKSKOEKSPKTAFFKQDTI	1617
DB	1362	NRPG-----RTPSKLLGP-----PMPK-EWESLEKSPKSTALTAKDII	1398
QY	1618	-LSLNACESNEAIAINEGONKPELEVTKAGRTERLCSONPPVLKHHQREITRTTQS	1676
DB	1399	SLPLDRHESNHSIAKNEGQATQREAAWTKQGGPRLCAPKPPVLRHHQKDIISLPTQOP	1458
QY	1677	DOEIDYDDT-SVEMKKEDFDIYDDEENQSPRSQKTRHYFIAAVERLWDYMGSSSPHV	1736
DB	1459	EEDKMDYDDIESTETKEDFDIYDDEENQSPRSQKTRHYFIAAVERLWDYMGSSSPHV	1518
QY	1737	LNRAQSGSVQPKVVPQETDQSTQPLVRGELNEHLGLLGYIRAEVEDNIMVTPRN	1796
DB	1519	LNRAQNGEVRFFKVVPRFADGSPQSYRGEINKHGLLGLGYIRAEVEDNIMVTPFN	1578
QY	1797	QASRPSYSSLIISVEEDQROGASPRKNFVXPNETKTYFWKVQHHMATEKDEFCKAWAY	1856

Db 1579 QASRPYSFSSLSYSPDDQSGAEPRHFNQPNETRYFVKVQHHMPTDEDDCKAWAY 1638
QY 1857 FSDVDLEKQVHSGVGLIGLLNCHNTLNPAHQGRQVTOEFPALFTTIDETKSWFTENNER 1916
Db 1639 FSDVDLEKQVHSGVGLIGLLNCHNTLNPAHQGRQVTOEFPALFTTIDETKSWFTENNER 1698
QY 1917 NCRAPCNQMEDPTFKENYRFFHAINGYIMDTPLGVNAQDQRIRWYLLSGMSENHSH 1976
Db 1699 NCRAPCHLQMEDPTLKENYRFFHAINGYIMDTPLGVNAQDQRIRWYLLSGMSENHSH 1759
QY 1977 FSGHVTVRKKEBKALYNLYFGVFTVEMLPKAGIWRVECLIGEHLAGMSTLFLVY 2036
Db 1759 FSGHVSVRKKEBKAVNLYFGVFTVEMLPKAGIWRVECLIGEHLAGMSTLFLVY 1818
QY 2037 SNKQCPPLGASGHIRDFQITAGQVQWAPKARLHYSGSINAWSTKPFHMKVDLLA 2096
Db 1819 SKKQAPLQVAGSRIRDFQITAGQVQWAPKARLHYSGSINAWSTKPFHMKVDLLA 1878
QY 2097 PMIHGKTQGAQKTSLSYISQFIIMYSIDGKKWOTYRGNSTGTLWVFFGNVDSGKIH 2156
Db 1879 PMIHGKTQGAQKTSLSYISQFIIMYSIDGKKWOTYRGNSTGTLWVFFGNVDSGKIH 1938
QY 2157 NFNPPILAIYIRLHTHYSIRTLRMELMGDLNCSMPGLWESKAISDAQTASSYFT 2216
Db 1939 NFNPPILAIYIRLHTHYSIRTLRMELMGDLNCSMPGLWESKAISDAQTASSYFT 1998
QY 2217 NFNATWSPKARLHLOGRNSAWPEQVNNPKWLOVDFQKTKMTGVTQGVKLLSMYV 2276
Db 1999 NFNATWSPKARLHLOGRNSAWPEQVNNPKWLOVDFQKTKMTGVTQGVKLLSMYV 2058
QY 2277 KFLISSSDGHQHTLFFQNGKVKVQGNQDSPTPVVNSLDPPLLYRIHRIHPOSVWHI 2336
Db 2059 KFLVSSQDGRWTLFLQDGHGKVKVQGNQDSPTPVVNSLDPPLLYRIHRIHPOSVWHI 2118
QY 2337 ALRMEVLGCAQDLY 2351
Db 2119 ALRLEVLCGAQDLY 2133

RESULT 4

KPHUS

coagulation factor V precursor [validated] - human
N:Alternate names: coagulation labile factor; proaccelerin
C:Species: Homo sapiens (man)
C>Date: 19-May-1989 #sequence revision 02-Jun-1995 #text_change 08-Dec-2000
C:Accession: A56172; A42344; A28028; A27498; A25837
R:Cripe, L.D.; Moore, K.D.; Kane, W.H.
Biochemistry 31, 3777-3785, 1992
A:Title: Structure of the gene for human coagulation factor V.
A:Reference number: A42344; MUID:92232668; PMID:1567832
A:Accession: A56172
A:Molecule type: DNA
A:Residues: 1-2224 <CRT>
A:Cross-references: GB:J03368
A:Accession: A42344
A:Molecule type: DNA
A:Residues: 48-58; 79-89; 120-130; 191-201; 238-249; 313-323; 368-378; 428-437; 461-471; 533-542;
2070; 2113-2120; 2172-2181 <CR2>
R:Jenny, R.J.; Pittman, D.D.; Toole, J.J.; Kriz, R.W.; Aldape, R.A.; Hewick, R.M.; Kaufm
Proc. Natl. Acad. Sci. U.S.A. 84, 4846-4850, 1987
A:Title: Complete cDNA and derived amino acid sequence of human factor V.
A:Reference number: A28028; MUID:87260886; PMID:3110773
A:Accession: A28028
A:Molecule type: mRNA
A:Residues: 1-857; 'R', 859-864, 'R', 866-924, 'E', 926-1763, 'I', 1765-2212, 'T', 2214-2224 <JEN>
A:Cross-references: GB:M18967
A:Note: Parts of this sequence, including the amino end of the mature protein, were deter
R:Kane, W.H.; Ichinose, A.; Hagen, F.S.; Davies, E.W.
Biochemistry 26, 6508-6514, 1987
A:Title: Cloning of cDNAs coding for the heavy chain region and connecting region of hum
A:Reference number: A27498; MUID:88107850; PMID:2827732
A:Accession: A27498

A:Molecule type: mRNA
A:Residues: 1-1284, 'I', 1286-1600 <KAK>
A:Cross-references: GB:M17785
A:Note: parts of this sequence were determined by protein sequencing
R:Kane, W.H.; Davies, E.W.
Proc. Natl. Acad. Sci. U.S.A. 83, 6800-6804, 1986
A:Title: Cloning of a cDNA coding for human factor V, a blood coagulation factor homolog
A:Reference number: A25897; MUID:86313685; PMID:3092220
A:Accession: A25897
A:Molecule type: mRNA
A:Residues: 1188-1215, 1315-2224 <KA2>
A:Cross-references: GB:M14335
A:Note: parts of this sequence were determined by protein sequencing
R:Keller, F.G.; Ortel, T.L.; Quinn-Alten, M.A.; Kane, W.H.
Biochemistry 34, 4118-4124, 1995
A:Title: Thrombin-catalyzed activation of recombinant human factor V.
A:Reference number: A56139; MUID:95210278; PMID:7696276
A:Contents: annotation; thrombin cleavage sites
C:Comment: Factor V is activated by thrombin and partially by coagulation factor Xa.
C:Genetics:
A:Gene: GDB:F5
A:Cross-references: GDB:119396; OMIM:227400
A:Map position: 1Q23-1Q23
A:Introns: 53/2; 84/1; 125/1; 196/1; 244/1; 318/1; 373/2; 432/3; 466/1; 537/3; 588/1; 65
C:Function:
A:Description: acts as a cofactor, with calcium and phospholipid, for the factor Xa prot
A:Pathway: blood coagulation
C:Superfamily: coagulation factor V; discoidin I amino-terminal homology; ferroxidase re
F:1-28/Domain: signal sequence #status predicted <SIG>
F:29-2224/Product: coagulation factor V #status predicted <MAT>
F:29-737/Product: coagulation factor Va heavy chain #status experimental <VAE>
F:29-345/Domain: A1 <DA1>
F:33-329/Domain: ferroxidase repeat homology <FO1>
F:346-691/Domain: A2 <DA2>
F:692-1573/Domain: B <DOB>
F:1183-1461/Region: 9-residue repeats (Q-X-T/N-L-S-P-D-L-S)
F:1574-2224/Product: coagulation factor Va light chain #status experimental <VAL>
F:1574-1905/Domain: A3 <DA3>
F:1667-1765/Region: ferroxidase repeat homology <FO3>
F:1906-2064/Domain: C1 <DC1>
F:1906-2224/Domain: C2 <DC2>
F:2065-2224/Domain: C3 <DC3>
F:2065-2221/Domain: discoidin I amino-terminal homology <DN2>
F:51-55, 239, 297, 460, 468, 554, 741, 752, 760, 776, 782, 821, 938, 977, 1074, 1083, 1103, 1106, 1479, 149
F:167-193, 248-329, 500-526, 603-684, 1725-1751, 1907-2061, 2066-2221/bisulfide bonds: #status
F:334-335/Cleavage site: Arg-Asn (protein C) #status predicted
F:363-693, 1546/Binding site: sulfate (Tyr) (covalent) #status predicted
F:376-377/Cleavage site: Arg-Ser (coagulation factor Xa) #status predicted
F:382, 1338/Binding site: carboxylate (Asn) (covalent) #status predicted
F:534-535/Cleavage site: Arg-Gly (protein C) #status predicted
F:737-738/Cleavage site: Arg-Ser (coagulation factor Xa, thrombin) #status experimental
F:1046-1047/Cleavage site: Arg-Trp (coagulation factor Xa, thrombin) #status experimental
F:1573-1574/Cleavage site: Arg-Ser (thrombin) #status experimental

Query Match 22.8%; Score 2824.5; DB 1; Length 2224;
Best Local Similarity 30.5%; Pred. No. 3e-139;
Matches 768; Conservative 390; Mismatches 850; Indels 507; Gaps 77;
QY 22 RRYLGAVELAWDMQSLGDLPLVDARPPRPVPSFPNTSVV-YKTLFVEFDHLFNI 80
Db 32 RQFVAAQGISWRPE-----PINSLSLSTVTFKTIIVREYDFY-FKK 75
QY 81 AKRPPPMGLLQPTQIAQVYDVTWITLKNMASHPVSLHAVGVSYMKAGEAYDDQTSOR 140
Db 76 EKQSTISGLLGLTYARVGLIKVFKKADKPLSHHPQIRYKSKSEGSYLDHTPPA 135
QY 141 EKEDKVPFGSGHYVWQVLKENGMSADPLCLTYSLYSHVDLVKDLNSGLIGALLVCRE 200
Db 136 EMDDAVAPGREGTYEWSISDSGTHDDPCLTHIYSHENLIEDFNSGLIGPLICKX 195

np000123-328-355-581.rpr

Wed Apr 14 15:52:00 2004

```
Db 1369 TVSPDLISLTLSPDPNEILAPDLQVTLSP-----DLIQTN----- 1405
Qy 1561 GKVPFLKAVATESAKTPSKLLDPLANDNHVGTQIPKEEMKSEKPEKTAFAKDKDTILSL 1620
Db 1406 --PALNHGHKASSADPDQASYPP--DSGQASSLP--ELNRTLPDPLTHIPPSPPSL 1458
Qy 1621 NACSENFALAINQGNKPEIEVTWAKQGRTERLCSONPPV--LKEHQBEIRITTLQSD 1677
Db 1459 N-----NTSLSKFNPLVVGLSRVGDGDEIVPSEE 1490
Qy 1678 QBEIDYDITISVEMKGDFOIYDECKQSPSSFOKKTRH-----YEI 1719
Db 1491 PERIDED-----YAEDDFVINYNDPYRDTTRDYNKSENPPDIAAWYLRHGCHGKKFYI 1544
Qy 1720 AAVRELNDYGMSSPHVLNRAQSGVQPQ---FKVVFQEFDTGDSFTQPLXGELNEHLG 1776
Db 1545 AAEETITWYAFQSEM--DHEDTGHIPKOTTYKVKVFKYLDSTFTSTEDPRAVEEHLG 1602
Qy 1777 ILGPVIRAEVDNIMVTRNOASRPSFYSSLIYSY-----EDQCGABPKRKNFKP 1828
Db 1603 ILGPVIRAEVDNIVQIRKPNLASRYSLSHAHGLSYEKSSEKTYDESPWFQEDAVQP 1662
Qy 1829 NETKTYFKVQVHMAFTKDEPCKAWAYPSVDVLEKDVHSLGICLLVCHTNTLPAHGR 1888
Db 1663 NSSYTVVHATKSGPENPGSACRAWAYYSANVVERDISHGLIGPLLCRKTLEMERL 1722
Qy 1889 QVTVQEPALPFTIFDTKSWYETENMERNCRAPCNIQMEDPTFKENYRPHAINGYIMDTL 1948
Db 1723 PDMREFVLLFMVDFBKKWYBKS--KGRS-----RISPEKNAHKFYAINGMIYN-L 1774
Qy 1949 PGVMAQDQIRWYLLSMGSENHISHFSGHVFTVRKKEEYKMYALNLYPGVFVEML 2008
Db 1775 PGLNTEQEWRLHLNMGSGSDIHVHPHGHQTLNRTKHQHLGVWPLPGSFKTEML 1834
Qy 2009 PSKAGIWRVECLJEGHLAGMSTLFLVSNKQTPGLMASCHISDFOITASGQYQWAPK 2068
Db 1835 ASKPGWLLDTEVGENQVAGQOTPFLLIDKECKRPMGLSTGVISDSQIKASELYITWEPR 1894
Qy 2069 LARLHYSGSINASTKE-----PFS-WIKVDLLAPMIHGIKTCQARQKPFSSLYISOFII 2122
Db 1895 LARLNNAGSNYNAWSEIKTALDPFKPIQVDMQKEVVVTGQTQAGXHYLKSCTFEQV 1954
Qy 2123 MYSLDGKKWQYRGNSTGTLVVFPGNVDSSGKGNINFPPIARIYIRLHPHYRSTLR 2182
Db 1955 AYSSDQTNQWLFGRSGKSVMYFTGNSDGTIKENRLDPPIVARIYIRHPTKSNRPTLR 2014
Qy 2183 YELMGCDLNSCMPLGMSKALSDAQITASSYFTNWEAT--WSPSKARLHLOGRNAMRPQ 2241
Db 2015 LELQCEVNGCSTPLUGLEDGRIQDKQITASSPKKSWGDYWEPSLARLNAQGRVNAWQAK 2074
Qy 2242 VNNPKENTQVDPQKTMKVGTGTQGVKSLLTMYVKFEILSSQDGHQWTLFFQNGKV-- 2299
Db 2075 ANNNKQWLOVDLLKIKVTAIVTQCKSLSSSEMYKSYISQYSDQGVAKPYRQKSSMVD 2134
Qy 2300 KVFQGRQDSFTPVVNSLDPPLITRYLRIHPQSNVQIARLMEVLGCEAODLY 2351
Db 2135 KTFEGSNTKGHMKNFNFPPIISRTIRIIPKTNQSIARLELFC---DIY 2183

RESULT 6
KF905
coagulation factor V precursor - bovine
C;Species: Bos primigenius taurus (cattle)
C;Date: 04-Mar-1993 #sequence_revision 28-Apr-1995 #text_change 11-Jun-1999
R;Accession: A42580; A36497
R;Giunto, E.R.; Esmont, C.T.; Mann, K.G.; MacGillivray, R.T.
J. Biol. Chem. 267, 2971-2978, 1992
A;Title: The complete cDNA sequence of bovine coagulation factor V.
A;Reference number: A42580; MUID:92147636; PMID:173753
A;Accession: A42580
A;Molecule type: mRNA
A;Residues: 1-2211 <GUI>
```

QY	320	MDLQQLSHCHSHQHDGMEAYVYVSCPEPOLIMKNEEAEYDDDDLTDSMDVVRP	379	Db	1195	QPSISPDGLQWALSPDPGQESLSPDLGQ	---	TSLSPDLSQESLSPDLGQTA	1242
Db	301	SPEGSWIASLPRHFQAGMAYIDIKNCAKTRPKK	---	LFRDQ	---	---	---	---	---
QY	380	DDNSPSFIQIRSVAKHPKTHVHIAAEEDWDYAPLVAPDDRSYKQSYNNKPGORIG	439	QY	1334	LKQFRLPLEBT	---	STLTQIDYNEKKGALITQ	1306
Db	344	---	---	Db	1243	LSPD	---	QESLSPDLGQALSPPQESLSPDLGQ	1294
QY	440	RKYKVRWYMTDETFR	---	QY	1387	SPISDCLTR	---	SHSIFQANRSPUPIAKVSFP9IRIYLTRVLFDQNSSHLPAASYRKK	1443
Db	369	KHYKVVYKQYQDDFTKALEDPSSGDLGPIIRAQVRDTLKIIVFNKMASRYSIYPH	448	Db	1295	TSLSPDLSQESLSPDLGQALSPPD	---	QESLSPDLGQALSPPD	1344
QY	498	GIT	---	QY	1444	DSGVQBSHLOGAKKNLSALITLMTGQBSLGTISATNSVYKVKVENTVLKPPD	1503	---	---
Db	449	GVTSPYDNEVNSSSTSGSNTMIRAVR	---	Db	1345	DLGOTSISPD	---	GOESLSPDLGQALSPPD	1381
QY	549	TRYSSFNWMEASLGLIPLLICVBSVDKGNQIMSDKNVILFSPDENRSMYITE	608	QY	1504	LPKTS	---	CKVILLPKVHIYQKDLFPETSNQSPGHLIDVEGSLLOGTGEGAIKNNEAN	1558
Db	501	TEPYNSVDITRLASGLIGLLIICAKSLDRGIQRAADISQAVFAVDENKSWYIED	560	Db	1382	LQOTSISPDLSQESLSP	---	DLGQALSPPD	1418
QY	609	NIQRELPNPAQVLEDPPEFQASNMH	---	QY	1559	RPGKVPFLRVATESSAKTPSKLLDPLAWDNHYGTQIPKEEWKSQESPEKTAFFKKKDTI	1617	---	---
Db	561	NIYKCEKPKVRODPKXVESNIMSNFTLPALINGVVPESI2ILGFCPDITVQMFCSVG	620	Db	1419	---	---	SPDLSQLDKQTSPP	1445
QY	663	AQTDPLSVFSGYTFKHVMVYEDTLTLPFSGETVFMSENPGLWILGCHNSDFNRGMT	722	QY	1618	---	---	NPVLKRXH	1665
Db	621	TQNDILTIHFTGSHFYGRHEDTLTFPMQGESVTVMQNVGTWMLITMNSNPRSKKL	680	Db	1447	PDLNQTSHTSOSSLSLPEPGQTFPNADIQOMSPSPDSLTNNTFPERBNPLVWVGL	1506	---	---
QY	723	ALLKVSCKNTGVYEDSYEDI	---	QY	1666	QRE	---	---	---
Db	681	LPRDAKLRNDD	---	Db	1507	SRDDGDYIBI	---	PROKESSEEDYGBFEFVAYNDPYQTLRT	1560
QY	756	IEPRFSQNSRHPSTQKQONATIP	---	QY	1708	---	---	FKVVFQFET	1758
Db	738	LGLRSPNSSL	---	Db	1561	WYLRSTGNRKYYITAAEISWDYS	---	FKVQDDVDYVPEDIVYKVKYKYL	1611
QY	815	QSTPHCLSLSDQEAKEVFSDPSGA	---	QY	1759	DSFTQYLRGELNEHGLIGVIRAEVDNIMVTFRNOASRPYSYSSLSIYE	---	---	1812
Db	784	RSHVRLIAKFAESLTLHLERAPAGSPLEHAGLDKNSALNP	---	Db	1612	DSFTKLDQGEVEHGLIGVIRAEVDNIMVTFRNOASRPYSYSSLSIYE	---	---	1671
QY	865	---	---	QY	1813	---	---	---	---
Db	839	DPREDHPLSDVTVGSLVLPFCGTNRKPAKQORFQVGRQAAKHKFSQTPPAKTRTL	898	Db	1672	TYEDDSEMPFKEDNAIQNKTYTYVWHAITRSGPENPGSACRAWAYYSANPEKDLSGL	1731	---	---
QY	905	ISTIPSDNLAAGTNTGS	---	QY	1871	IGPLLCHVNTLNPAGHQVTVQBPALFTTIPDETSTWYPTENNERNCRAPCNIOMEDPT	1930	---	---
Db	899	---	---	Db	1732	IGPLLCRKTLCKENMPVDMMREFVLLFVFPDEKSWYDVKKPTRSWRASS	---	---	1785
QY	955	LSLSEENNDKLSGLMNSQES	---	QY	1931	FKENYRPHALNGYIMDPLQVLMAQDQIRWYLLSMGNSNIHSIHPSGHVTVTKKEEY	1990	---	---
Db	943	YEITQDANENTVVK	---	Db	1786	VKNSHEFHAIINGMYN	---	---	---
QY	1005	---	---	QY	1991	KMALNLYPGVFTVEMLPKAGIWKVVECLIGELHAGMSTLFLVYKNCQTLPLGMASGH	2050	---	---
Db	998	DRQARNRSLKEGLLR	---	Db	1845	QLGWPLLPSPFKTEMLKASKPGWLLDTEVEIQORAGMQTPFLIVDRECKMPGLSTGL	1904	---	---
QY	1055	DTEFKVVTPLIHDRMLMDKNATALRLNMSKNTTSSKNMVMQKKEGPIEPDAQNPMWS	1114	QY	2051	IRDPQITASGOYQOWAKLAELHYSGSINAW	---	---	---
Db	1049	DRHN	---	Db	1905	IADSIQIASEFWGWEPLAKRLNKGSGYNAMAEKLSFEPFEPWIOVDMQKEVLLTGICQ	1954	---	---
QY	1115	FFQMLFLESARWIORHCKNSLN	---	QY	2105	TOGARQKPSYISQFIIMYSIDGKQWQTVGNSGTGLMVFNGVSDSGIKHNFNPII	2164	---	---
Db	1088	---	---	Db	1965	TOGAKHYLKPYTTEFCVAYSLDRKWRI	---	---	---
QY	1172	GRGEFTKDYGLKEMVPPSSNNIPLTNLDNLHNNHTNQEKQOBEIEKKEITLIQENVLP	1231	QY	2165	ARYRLHRTHTVSIKSTLRMBLNGCDLASCMLPGMESKAI	---	---	---
Db	1118	---	---	Db	2025	ARYIRISPTGVSFKPALRELOQCEVNGCSTPLNGESKIKENKQITASSPKSWGNWE	2084	---	---
QY	1232	QIHVTGTGNFMKNMFLSTRQNEGSYD	---	QY	2224	PSKABLHLOGRKNAPQVNNPKELQYDFOKTKMTVGTQGVKSLTSMYVKEFLSS	2283	---	---
Db	1140	PTHSTTAPSRS	---	Db	2085	PFLAELNAQGRVNAWAKANNNNQWLQDLKIKITAIIVTOGCKSLSESMYKSYTHY	2144	---	---
QY	1274	NDSTNRKGTAFHFKGBEENLEGLNQKQIVKAC	---	QY	2284	SODGHQWTLFFQNGKV	---	---	---

Db 2145 SDQGTWKPYREKSSMVDKIFEGNNVRGVHVFNPFPILSRFIRIIPKTKWQSIARLE 2204

Qy 2342 VLGCFAQDLY 2351

Db 2205 LFQC---DMY 2211

RESULT 7

A25945

coagulation factor VIII - pig (fragment)

C:Species: Sus scrofa domestica (domestic pig)

C:Date: 16-Aug-1988 #sequence_revision 16-Aug-1988 #text_change 21-Jul-2000

C:Accession: A25945

R:Toole, J.J.; Pittman, D.D.; Orr, E.C.; Murttha, P.; Wasley, L.C.; Kaufman, R.J.

Proc. Natl. Acad. Sci. U.S.A. 83, 5939-5942, 1986

A:Title: A large region (approx 195 kDa) of human factor VIII is dispensable for in vitro

A:Reference number: A25945; PMID:86287369; PMID:3016730

A:Accession: A25945

A:Status: nucleic acid sequence not shown; not compared with conceptual translation

A:Molecule type: DNA

A:Residues: 1-869 <TOO>

C:Superfamily: coagulation factor VIII; discoidin I amino-terminal homology; ferroxidase

Query Match 17.6%; Score 2190; D3 2; Length 869;

Best Local Similarity 45.2%; Pred. No. 1.1e-106;

Matches 493; Conservative 122; Mismatches 251; Indels 224; Gaps 15;

Qy 705 GLWILGCHNSDFRNRGMTALLKVSSCDKNTGYEDSYSDISAYLLSKNNAIPRFSQN 764

Db 1 GLWVLGCHNSDFRNRGMTALLKVSCDRTGDYDNTYEDIPGFLLSGKNVIEPRSPAQN 60

Qy 765 SRHPSTROKFNATIPENDIEKTDWFAHRTMPKIQNVSSDDLMLLRQSTPHGLSL 824

Db 61 SRPPSASQOFITSPEDDVE-LDPQSQRTOALELSVPSDGGSKMLLQGNPAPHGSS 119

Qy 825 SDIOEAKYETSPDGPATDSNLSSEMTFRPQLHSGDMVFTPESGIQLRLNEKLG 884

Db 120 SDIOEARNE--ADYLPGARERNVTAASAAARLPELHHSARVLTETP-----EK--- 167

Qy 885 TAATEKGLDFKVSSTNNLIS--TIPSDNLAAGTNTSSLGPPSPVHVDSQIDTLFLG 942

Db 168 ----ELKLDKXSSSDLLKTSPTIPSDLSAETERTHSLGPPHPQVNFPSQIGALVIG 223

Qy 943 KKSPLTESGGLSLSENNDSKLLSGIMNSQESSWGNKVSSTESGLPKGKRAHGAPAL 1002

Db 224 KNSSHFTGAGVPLGSTEDH-----ESSLGENVSPVESDGIPEKRAHGAPAS 270

Qy 1003 LTKDNALFKVISLTKNTKSNNSATNRKTHIDGPSLLIENSPPVWQNLLESDETFPKVT 1062

Db 271 LTKDDVLFKNISLVKTKKARVYLKTRKIHIDDAALLTENRASA----- 325

Qy 1063 PLIHDRMLKDNATALRLNENSNKTTSSKNMVMQKKEGIPDPAQNPMSFFKMLPLP 1122

Db 316 -----TFMDKNTTASGLNHVSN----- 332

Qy 1123 ESARKIORTGKNSLNSQSGSKPOLVSLGPEKSEVGNFVLSKKNVVGKGETKDVGL 1182

Db 333 ---WIKGPKNPLSBERGSEPELLTSSGSKSVKQSGQGRIRVAVEEELSKG--- 385

Qy 1183 KEMVFPSSRNPLTNLDNLNHNTHNQEKIQQEIEKKTILQENVPLPQIHTVTGKNF 1242

Db 386 KEMMLPNSLPLTNSADVQNDTHSQGKSEEMEREKLVQPKVDLPQVVTATGKNF 445

Qy 1243 MKNLFLSLTRQNEGSDGAYAPLVQDFRSLNDSNTNRKTHTAHPSKKGREENLGLGNQ 1302

Db 446 LRNIFHQSTESVEGSDGSHAPVQDGRSLNDSAEAEATHAFSAIREAPLEAPGNF 505

Qy 1303 TKQIVEKVACTRISPTNSQNFVQTSKRAKQFRLPLETELEKRIIVDDTSQWSEN 1362

Db 506 T-----GPGPSAVPRVRKQSLKQRLPLEEKIPRGVVLNATIRWS-- 548

Qy 1363 MGLHPTSLTQIDYNEKEKGAITQSPSLDCLTRSHSIPOANRSPLPKAVSSFPISRTY 1422

Db 549 ----- 548

Qy 1423 LTRVLPDQNSHLLPAASVRKKKSGVQESSHFLOGAKKNNLSLALLTLEMTGDQREVGSGL 1452

Db 549 -----ESSPLQGAKNLNLSPFLTLEMAAGGQKISALG 582

Qy 1483 TSATNSVYKKVENTVLPKPLPKTSQKVELLPKHVYQKDLFFETTSNGSPGHLDLVEG 1542

Db 583 KSAAGPLASGKLEKAVLSSAGLSEASGKAEPLPKVRVEREDLLPKTSNVSCHAGDLGQE 642

Qy 1543 SLLQGTGALKWNEANRPCKVPFLRVATESAKTPSKLLDPLANDNHVGTQIPKEEWSQ 1602

Db 643 IFLOKTRGFVNLRKVRP-----RTPSKULGP-----PMK-EWESL 679

Qy 1603 EKSPEKTAFAKKDII-LSLNACSNHAAINEGONKPEIEVTWAKQRTRELCSQNPVP 1661

Db 680 EKSPASTALRTKDIISLPLDRHESNHSIAAKNEGOAETQREAAWTKQGPQRLGAPKPPV 739

Qy 1662 LRRHQREITRTTQSQDQBEIDYDITISVEMKKEDPDIIVDEBNQSPRSFQKTRHYFTAA 1721

Db 740 LERHQDILSLPTROPEEDKMDYDDIFSTETKGEDEFIYGEDENQDPRSFQKTRHYFTAA 799

Qy 1722 VERLDYGMSSSPHVLNRPAQSGSVPOKQKVVQFETDGSFTQPLRYGELNEHLGLGPF 1781

Db 800 VEQLDYGMSSSPALNRPAQNGEVRPKKVVVRBRADGSGFTNPSYRGELNKLGLGPF 859

Qy 1782 IRAEVEDNIM 1791

Db 860 IRAEVEDNIM 869

RESULT 8

KUHU

ferroxidase (EC 1.16.3.1) precursor [validated] - human

N;Alternate names: ceruloplasmin

C;Contains: ferroxidase long form (CP-1); ferroxidase short form (CP-2)

C;Species: Homo sapiens (man)

C;Date: 31-Aug-1980 #sequence_revision 12-May-1995 #text_change 08-Dec-2000

C;Accession: A25443; A25445; A35450; A00524; I59067

R:Koshchinsky, M.L.; Funk, W.D.; van Coest, B.A.; MacGillivray, R.T.A.

Proc. Natl. Acad. Sci. U.S.A. 83, 5086-5090, 1986

A:Title: Complete cDNA sequence of human preceruloplasmin.

A:Reference number: A25443; PMID:86259737; PMID:2873574

A:Accession: A25443

A:Molecule type: mRNA

A:Residues: 1-1060,1065-1069 <KOS>

A:Cross-references: GB:M13699; NID:G180255; PIDN:AAA51976.1; PID:g-80256

A:Note: this is the short or CP-2 alternatively spliced form

R:Mercer, J.F.B.; Grimes, A.

FEBS Lett. 203, 185-190, 1986

A:Title: Isolation of a human ceruloplasmin cDNA clone that includes the N-terminal lead

A:Reference number: A24165; PMID:86275241; PMID:3755405

A:Accession: A24165

A:Molecule type: mRNA

A:Residues: 1-401549-599;784-829;919-952 <MBE>

R:Yang, F.; Friedrichs, W.E.; Cupples, R.L.; Bonifacio, M.J.; Sanford, J.A.; Horton, W.A.

J. Biol. Chem. 265, 10780-10785, 1990

A:Title: Human ceruloplasmin. Tissue-specific expression of transcripts produced by alte-

A:Reference number: A35450; PMID:90285218; PMID:2355023

A:Accession: A35450

A:Molecule type: DNA

A:Residues: 1007-1064 <YAN>

A:Cross-references: GB:J05506

A:Note: this is the long or CP-1 alternatively spliced form

R:Takahashi, N.; Ortel, T.L.; Putnam, F.W.

Proc. Natl. Acad. Sci. U.S.A. 81, 390-394, 1984

A:Title: Single-chain structure of human ceruloplasmin: the complete amino acid sequence

A:Reference number: A00524; PMID:84119493; PMID:6582496

A:Accession: A00524

A:Molecule type: protein

A:Residues: 20-1060,1065-1069 <YAK>

A:Note: 79-Gly and 449-Gly were also found

R:Yang, F.; Naylor, S.L.; Lum, J.B.; Cutshaw, S.; McCombs, J.L.; Naberhaus, K.H.; McGill

Proc. Natl. Acad. Sci. U.S.A. 83, 3257-3261, 1986
A>Title: Characterization, mapping, and expression of the human ceruloplasmin gene.
A'Reference number: 159067; MUID:86205876; PMID:3486416
A'Accession: 159067
A>Status: translated from GB/EMBL/DBJ
A'Molecule type: mRNA
A'Residuals: 218-1059 <RES>
A'CROSS-references: GB:M13536; MID:g180248; PIDN:AA51975.1; PID:g180249
C'Comment: Ferroxidase is a blue, plasma alpha2-glycoprotein binding 6-7 copper ions per molecule. In Wilson's disease the plasma levels of ferroxidase are diminished or undetectable. The three fragment chains are produced spontaneously during purification and are not artifacts.
C'Keywords: acute phase; alternative splicing; copper; duplication; glycoprotein; oxidoreductase; signal sequence; status predicted <MATL>
F:1-19/Domain: signal sequence; status predicted <MATL>
F:20-1069/Product: ferroxidase, long form #status predicted <MATL>
F:20-1060/Product: ferroxidase, short form #status experimental <MATL>
F:20-499/Product: ferroxidase 67K chain #status experimental <K67>
F:23-357/Domain: ferroxidase repeat homology <F01>
F:373-718/Domain: ferroxidase repeat homology <F02>
F:501-905/Product: ferroxidase 50K chain #status experimental <K50>
F:733-1059/Domain: ferroxidase repeat homology <F03>
F:907-1065/Product: ferroxidase 19K chain #status experimental <K19>
F:138,397,762/Binding site: carboxylate (Asn) (covalent) #status experimental
F:174-200,276-357,534-560,637-718,874-900/Disulfide bonds: #status predicted
F:227,588,926/Binding site: carboxylate (Asn) (covalent) #status absent
F:235,338,343/Binding site: copper (His, Cys, His) (type 1) #status predicted
F:338/Binding site: carboxylate (Asn) (covalent) (partial) #status experimental
F:656,699,704,709/Binding site: copper (His, Cys, His, Met) (type 1) #status predicted
F:994,1040,1045,1050/Binding site: copper (His, Cys, His, Met) (type 1) #status predicted

Query Match 10.7%; Score 1326; DB 1; Length 1069;
Best Local Similarity 19.2%; Pred. No. 2.6e-61;
Matches 399; Conservative 204; Mismatches 419; Indels 1054; Gaps 29;
Qy 5 LSTOFFCLLPSCATRRYVLGAVELSDWYQSDLGE--LPVDARPPPRVPSFPENT 61
Db 6 LGIFLFLCSTP-AWAKEHYVIGIEITWDY-ASDHGSKLISVDTEHSNTYLONGPDRI 63
Qy 62 SVVYKTLFVETDLENIAPRPDWMGLGFTIOAEVYDVTWITLKNMASHPVSLHAGV 121
Db 64 GRLYKALYLQTDTEFTTEKPVWLGFLGPIIAETGDKVYVHLKLAGRPYTFHSHG 123
Qy 122 VSYWKASGAEYDDQTSOREKEDDKVFGGSHYVQVYKENGPMASDPLCLITYLSHV 181
Db 124 ITYYKEHGAIPYDNTDFQADRVYQEQYTYMLLATEEQSGEGGNCVTRIYHSHI 183
Qy 182 DLVQNLNGLICALLVCEGSLAKETQTL-HKFILLFAVDECKSWHSE-----TKNS 234
Db 184 DAPKDIASGLIPLICKDSADKEKHIHIDREFVMSVDENFSDYEDNIKTYCSEP 243
Qy 235 LMQDDASARAWPKMHTVGVNRSILPGLIGHCRKSVYVHVGITTPVHVSIFLEHT 294
Db 244 EKVDKDNEDFQSNRMYSVNGTTFGLSLPGLSNCAEDRVKWLFGMGNEVDVHAFFHQA 303
Qy 295 FLVNRHQASLEISPIITFTLAQTLMDLQGLFSLCHSHQHDGMEAYVKVDSCEEPQL 354
Db 304 LINKVRIITNLFPATLDAYVAQNPGEWMLSCQNLNEKLAGLQAPFQVQEC----- 357
Qy 355 YKQNEEAEYDDDLTDSMDVVRFDNPSPSFIQIRSVAKHPKPTWVHIAEEEDWDY 414
Db 358 ---NKSSKD-----NIRKXVH-----YIAAEELIWN 385
Qy 415 APL-----VLAPDRSYKQYLNNNGPQIRGRYKVRFMAYTDTF---KTREAIC 462

386 APSGIDIFTKENLPAGSDS--AVFFQGTTRIGSGYKLVYREYTDASFNNRERGPBE 443
Qy 463 HESGILGPLYGEGDITLLIIFKNOASRPYNIYPHGI-----TDVRPLY---SRRLP 511
Db 444 EHLGILGIVIAEYVGGDITIRVTFHNGAYPLSIEIGVRFNKNNEGTYSPNPNQSRSPV 503
Qy 512 KGVHKLDFPILPGEIFKYKWTVTVEDGPTKSDPRCLTRYYSFVNMRERDLASGLIGPLL 571
Db 504 PSASH-----VAPTEFTYEWTVPEVGTNADPVCLAKTYSAVDPTKDTFTGLIGPMK 558
Qy 572 ICYKESVDCKGNQMSDKRNVLISVFEDENSWLTENIQRFELPNPAGVQLEDDEFOASN 631
Db 559 ICKGSLHANGREQDVKDFEYLFPTVFDENESLLEDNIRMTTAPDQVDEDEFOASN 618
Qy 632 IMGSHGVYVDSLQ--LSVCLHEVAYWYLSICAQTDFLSFYFSGYTFKHKMYEYDTLTF 690
Db 619 KXHSWNGFWYGNQPGLTWCKGDSVWYVLFSAQNEADVHGIYFSGNLYLWGRERRDTANLF 678
Qy 691 PPSGETVFMSENPGLWILGCHNSDFNRGMTALLKVSQCDXNTGTYEDSDYEDSAYILL 750
Db 673 POTSLLHWPDECTENVECLTTHYTGGMKQKVTNQ------- 718
Qy 751 SKNAIEPRSFQNSRHSRSTROKQVATTPENDIEKDPWFHRTPEMPKIQNVSSDILL 810
Db 719 ---PRQSE-----DSTFYLGYTY-- 735
Qy 811 MLLRQSPTPHGLSLDLQAKYETSDDPSPGAIDSNNSLSBMTFRPOLHHSQGMVFTP 870
Db 736 ---IAAVEVE----- 742
Qy 931 HYDSQLDITLFGKSSPLTESGGPLSEBENNDKLESLMNSQESSWGKNVSTESGR 990
Db 743 --- 742
Qy 991 LFKGKRAHPALLTKDNALFKVSIKTKNTKNSATNRKTHIDGFSLLIENSVPWQN 1050
Db 743 ---WD- 744
Qy 1051 ILSDTEFKKVTFLIHDRMLMDKNAFALRNHMSKNTSSKNMEMVQKKEGPIPPDAQN 1110
Db 745 --- 744
Qy 1111 PDKSFFKWLFLPESARWIQTHGKNSLNSGGPSPKOLVSLGPEKSVGQNFLEKKNV 1170
Db 745 --- 744
Qy 1171 VGKGETRQVLEKEMVFPSSRNLFITNLNLHENNTNQEKKIOEBIEKKEKTLIQENVVL 1230
Db 745 --- 744
Qy 1231 PQIHVTGTKNFMKNLFLSTRQNVGSDGAVAPVLDQFSLNDSTNRKTKHTAHSKK 1290
Db 745 ---YSP----- 747
Qy 1291 GEEENLEGLNQTKQIVEKYACTTRISNTSQNFVQTSKRAKQFELPLETELEKRI 1350
Db 748 ---OR----- 749
Qy 1351 IVDDTSTOWSKMKHLTFTLTQIDYNEKEGATQSPSLDCLTRSHSIPQNSRPLPIA 1410
Db 750 ---EWEKELHHL----- 758
Qy 1411 KVSPPSPRIPIYLTREVLFDQNSSHLPAASYSKKSQVQESSHFLQAKKNLSLAILTLE 1470
Db 759 --- 768
Qy 1471 MTGQDQREVSLGTSATNSVYKVKVENTVLPKLPKTSKGVKELLPKVHIYQKDLPTETS 1530

Arch. Biochem. Biophys. 293, 1-8, 1992
A;Title: Rat ceruloplasmin: resistance to proteolysis and kinetic comparison with human
A;Reference number: S21692; MUID:92117681; PMID:1531003
A;Accession: S21692
A;Molecule type: protein
A;Residues: 20-25; Q; 902-910 <RYA>
C;Superfamily: ferroxidase; ferroxidase repeat homology
C;Keywords: copper; glycoprotein; oxidoreductase; plasma
P;1-19/Domain: signal sequence #status predicted <SIG>
P;20-1059/Product: ferroxidase #status predicted <MAT>
P;23-356/Domain: ferroxidase repeat homology <FO1>
P;372-712/Domain: ferroxidase repeat homology <FO2>
P;727-1053/Domain: ferroxidase repeat homology <FO3>

Query Match 10.3%; Score 1277; DB 1; Length 1059;
Best Local Similarity 19.1%; Pred. No. 9.3e-59;
Matches 395; Conservative 196; Mismatches 424; Indels 1054; Gaps 29;

Qy 5 LSTCFFPLCLRFPQSAATRRYILGAVELSWDMQ-SDLGEL-PVDARFPVRPKSEFNTS 62
Db LSALLFL-HSSLAWTREKHVYIGITEAVWDYASSEEKELISVDTOSNFVLRNGPDRIG 64
Qy 63 VVYKTLFVEFTDHLNIAKPRPMMGLGPTIOAEVDTVITLKNWASHPVLSHVG 122
Db 65 RYKXKALYSSTDTGTFKTDKPAWLGFLGPIKAEVCDKYSVHVKNFASRYTFHAGV 124
Qy 123 SYWKASGAZYDDTSQREKEDKVPFGGSHYVQVVKENGPMASDPLCLTYLSHVD 182
Db 125 TYTKANEGALYDNTDFQADKLFQOQYLYLRA-NEPSPGEGDSNCTRIYHSHVD 183
Qy 183 LVKQNSGLIGALLVCRGSLAKETQTL-HKFTLLFAVFEDEGKSHSETKNSLM----- 236
Db 184 APKDIASGLIGLICKGSLHKEENIDQEFVLMFSVVDENLSWYLEDNIKTFCSBPE 243
Qy 237 -QDRDAASARAWPKXHTVNGVNRSLPLGLIGHRKSVYHVMIGMTTPEVHSIFEGHTF 295
Db 244 KVDKQNEDEPQESNEMYSINGYTFGLSLGSMCAEDRVAKYVLFMGNEVDVHSALFHGOAL 303
Qy 296 LVRNHRQASLEISPIITFLTAQTLMDLQGLFISHSQHDGMEAYKVQDSCPEEPQLT 355
Db 304 TSKNYHTDIINLFPATLIDVSMVAQNPQVWMLSQNLNHLKAGLQAFQVRDC----- 356
Qy 356 MKNNEAEDYDDLTDSEMDVVRFDNDSFSPFIQIRSAKHPKTHVHYIAAEEDWDYA 415
Db 357 --NKPSP--DDDIQDRHV-----RH-----YIAAEITWDYA 385
Qy 416 P-----LVLAPDORSYKQYLNNGPORIGRKYKVRFMAYTDTF---KTRSAI 461
Db 386 PGGTDTFTGENTSLGSDSRVFEQ---GATKIGGSKYKLVREYTDSDFTNRKERGPD 441
Qy 462 QHESGILGLLYGEVGDITLIIFKQNASRPNYIYHGHITDVR-----PLYSRLPKGVKHL 517
Db 442 EEHLGILGPVIAEAVGDIIRVTFNKQFPFLGIQPMGVFTKENEYTYG---PDGRSK 498
Qy 518 KDFPILPGRIFPKYKMTVTVEDGPTKSDPRCLTRYSSFVNMRDLASGLIGPLICYKES 577
Db 499 QASHVAPKETFTYVETVTPKEMGTATYAPVCLSKMYISGVDLTKDITFTGLIGPKICKGS 558
Qy 578 VDQKQINMSDKENILFSVDENRNSWYLTENFORFLPNAGVQLEDPEPQASNIHSHIN 637
Db 559 LLADGRQKVDKFFYLPATVFDENESLLDDNTRMTAPENVDKEDDEPQESNKHSMN 618
Qy 638 GYVDSIQ-LSVCLHEVAYWYILSIGAQDTFLSVFSGYTFKHKVYVEDTLTLFPSPSET 696
Db 619 GFWYGNLPGLNMLGIESIVWYLFSAAGNEADHGHYFSGNTYLSKGERDRTANLFPKSLT 678
Qy 697 VFMSMENPGLWILGCHNSDFRNGMTALLKVSCKDNTGVDYEDSDYBDISAYILLSKNNAI 756
Db 679 LLMTPTDEGDFVDECLTDTHTGTGMKQYTVNQC-----KQGFEDVT----- 720
Qy 757 EPRFSQNSRSEPTQKQFNATTIPENDIEKTDPEFAHRTMPKIQNVSSDILLMLRQS 816
Db 721 ----- 720

1531 NGSPGELDLVEGSLLOGTGAIKRWNEANRPKVPFLRVATPESAKTSKLLDLPAWNNHY 1590
769 ----- 768
1591 GTQIPKEWKSQEKSPKTAFAKKDDTILSNACSNHAIKALNEGQNKPSIEVTKAQGR 1650
769 ----- 768
1651 TERLCSQNPVLKHFQREITRTTLQSQDEIDYDDTISVEMKXEDFDIYDEDENQSPRSF 1710
769 ----- 772
1711 QKTRHYPFAVERLWDYGMSSPHVLNRNRAQSGSVQFKVFOEFTDGSFTOPLVRGE 1770
773 -----FVIGS-----KYKVVYRQVYTDSTFRVPEVKA 800
1771 LNEHLGLGPVIRAEVENIIVTRNQASRPSYFSSLSISEEDORQAERKPFVK--- 1827
801 EEEHLGLGPQLHADVGDKVAKIIFKNMATRPSYIHA-----HGVQTESSTVITPTL 850
1828 PNETKTYEYKQHWMAPTKQREDFCKAWAYFGDVLKEDVHSLGLIGPLLVCHTNTINPAHG 1887
851 PGETTYVWKIPERSGAGTESACIPNAYYSIVDQVKLYSLIGPLIVCRPYLKVENP 910
1888 RQVTVQVEFALFTTIFDETKSYFTYFTENMERNCRAPNIQMEBPTFKENYRFAINGYIMDT 1947
911 RRKL--EFALLFLVFDENESWYLDNIIKTYSDPEKVKDKDEFTESNKHAINRGMFGN 968
1948 LPGLVMAQDQIRWILLSGNSNENHSHFSGHVTTRVKEEYKMAKLYNLPVGPETVEM 2007
969 LQGLTMHVGDENVWYLMNGNIDILHTVHFHGHSGYQYKRGYSSVDVDFIPFTYQTLLEM 1028
2008 LPSKAGIWRVEGLICEHLHAGMSTILFLVYSNKCQTP 2043
1029 PFRFTGILLHCHVTDHIIHAGMETTVTLVQLNEGEYP 1064

RESULT 9
A35210
ferroxidase (EC 1.16.3.1) precursor - rat
N:Alternate names: ceruloplasmin
C:Species: Rattus norvegicus (Norway rat)
C:Date: 10-Sep-1999 #sequence revision 10-Sep-1999 #text change 10-Sep-1999
C:Accession: A35210; A41753; A29564; S21692
R:Fleming, R.E.; Gitlin, J.D.
J: Biol. Chem. 265, 7701-7707, 1990
A:Title: Primary structure of rat ceruloplasmin and analysis of tissue-specific gene exp
A:Reference number: A35210; MUID:90237081; PMID:2332446
A>Status: not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 1-1059 <FL2>
A:Cross-references: GB:J05424
R:Fleming, R.E.; Gitlin, J.D.
J: Biol. Chem. 267, 479-486, 1992
A:Title: Structural and functional analysis of the 5'-flanking region of the rat cerulop
A:Reference number: A41753; MUID:92112697; PMID:1730611
A:Accession: A41753
A:Molecule type: DNA
A:Residues: 1-48 <FL2>
A>Note: the authors translated the codon GAA for residue 40 as Gly, GAA for residue 41 a
R:Alfred, A.R.; Grimes, A.; Schreiber, G.; Mercer, J.F.B.
J: Biol. Chem. 262, 2875-2878, 1987
A:Title: Rat ceruloplasmin. Molecular cloning and gene expression in liver, choroid plex
A:Reference number: A29564; MUID:87137545; PMID:3818625
A:Accession: A29564
A:Molecule type: mRNA
A:Residues: 'NSG', 215-216, 'Y', 218, 'FAT', 222, 'F', 224-226, 'E', 228, 'LL', 231, 'D', 233-235, 'RM
A:Experimental source: liver
A>Note: the authors translated the codon GCG for residue 60 as Gly and GTG for residue 1
R:Ryan, T.P.; Grover, T.A.; Aust, S.D.

QY	1891	TVQEFALFTTIFDETSKSWYFTENMRNCRAPCNIQMEDPTFKKNYRFAHNGVIMDTLPG	1950
Db	906	KMEFSLLFLVFDENESWYLDNNINTYDPHPKVKNDNEEFIESNKKVHAINGFQVGNLQG	965
QY	1951	LVAQDQRIWYLLSGNSNENIHSIHSGVFTVRKKEEYKALYNLYPGVSTVEMLPS	2010
Db	966	LTHVGDENVWYVMAGNFEIDLETVHFHGHSGFYQKHGHISSDVFDFPFGTYQIEMFPQ	1025
QY	2011	KAGIWRVECLIGRHLHAGMSTLFLVYSNK	2039
Db	1026	TPGWLHLCHVTDTHIHAGVTTYTVLPNQ	1054
RESULT 10			
factor VIII-associated gene B hypothetical protein - human			
C/Species: Homo sapiens (man)			
C/Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 18-Jun-1999			
C/Accession: A44258			
R:Levinson, B.; Kenwick, S.; Camel, P.; Fisher, K.; Gitschier, J.			
Genomics 14, 585-589, 1992			
A/Title: Evidence for a third transcript from the human factor VIII gene.			
A/Reference number: A44258; MUID:93052386; PMID:1427887			
A/Accession: A44258			
A>Status: preliminary			
A/Molecule type: mRNA			
A/Residues: 1-216 <LEV>			
A/Cross-references: GS:M9C707; NID:G182316; PID:AAA58466.1; PID:G182317			
C/Superfamily: coagulation factor VIII; discoidin I amino-terminal homology; F			
F.1-53/Domain: discoidin I amino-terminal homology #status atypical <DN1>			
F.57-210/Domain: discoidin I amino-terminal homology <DN2>			
Query Match 8.9%; Score 1104; DB 2; Length 216;			
Best Local Similarity 100.0%; Pred. No. 7.9e-51;			
Matches 208; Conservative 0; Mismatches 0; Indels 0; Gaps 0			
QY	2144	VFFGNVDSSGKHNIFNPITARYIRLHPHTYSTRTERMLMGCDLNSCSMPLGWSKA	2203
Db	9	VFFGNVDSSGKHNIFNPITARYIRLHPHTYSTRTERMLMGCDLNSCSMPLGWSKA	68
QY	2204	ISDAQITASSYPTNFWATWSFKARLHLQGRSNARPPQNNPKWLQVDFQKTMKVTGVT	2266
Db	69	ISDAQITASSYPTNFWATWSFKARLHLQGRSNARPPQNNPKWLQVDFQKTMKVTGVT	128
QY	2264	TGQVKSLLTSMVVKFLISSSDGQHWTLFFQNGKVKVFGQNQDSFTPVNSLDPLLLTR	2322
Db	129	TGQVKSLLTSMVVKFLISSSDGQHWTLFFQNGKVKVFGQNQDSFTPVNSLDPLLLTR	188
QY	2324	YLRIHPQSWVHOIALRMEVLGCEAODLY	2351
Db	189	YLRIHPQSWVHOIALRMEVLGCEAODLY	216
RESULT 11			
Jc4915			
ags protein precursor - rat			
N:Alternate names: O-acetyl-Gd3 ganglioside			
C/Species: Rattus norvegicus (Norway rat)			
C/Date: 26-Sep-1996 #sequence_revision 01-Nov-1996 #text_change 20-Jun-2000			
C/Accession: Jc4915			
R:Ogura, K.; Nara, K.; Watanabe, Y.; Kohno, K.; Tai, T.; Sanai, Y.			
Biochem. Biophys. Res. Commun. 225, 932-938, 1996			
A/Title: Cloning and expression of cDNA for O-acetylation of GD3 ganglioside.			
A/Reference number: Jc4915; MUID:96374422; PMID:8780713			
A/Accession: Jc4915			
A>Status: preliminary			
A/Molecule type: mRNA			
A/Residues: 1-427 <OCU>			
A/Cross-references: DDBJ:D84068; NID:G1620006; PIDN:EAA12210.1; PID:G1620007			
A/Experimental source: CST cell			
C/Comment: This protein is required for the O-acetylation of disialoganglioside			
C/Genetics:			
A/Gene: ags			

C;Superfamily: milk fat globule protein; discoidin I amino-terminal homology; EGF homol
F;1-21/Domain: signal sequence #status predicted <SIG>
F;28-60/Domain: EGF homology <EG1>
F;68-107/Domain: EGF homology <EG2>
F;110-267/Domain: discoidin I amino-terminal homology <DN1>
F;271-427/Domain: discoidin I amino-terminal homology <DN2>

Query Match 5.3%; Score 663; DB 2; Length 427;
Best Local Similarity 40.3%; Pred. No. 2.9e-27;
Matches 142; Conservative 61; Mismatches 123; Indels 26; Gaps 6;

Qy 2018 ECLIGELHAGMSTLFLVYSNK-----COTPLGASGHIRDFQITASQY-- 2062
Db 78 KCLVTEIQRG--DIFTEYICQCPGVSGIHCEIGCCSKLGLGGAIADSOISASSVMG 135

Qy 2063 ----GOWAPKLARLHYSGSINAW--STKEPSWIKVDLLAPMIHGIKTQAGKFFSLY 2116
Db 136 FMGLQRWGPELARLYRTGIVNANTASSYDSDKFWIQVDFLRKRVSGVMTCQASAGRAEY 195

Qy 2117 ISQFLIMISLDGKKQWYRGNSTGLMVFVFGVDSGGIKHNIENPPIIARYIRLHPHYS 2176
Db 196 LKTEKVAYSLDGRPEFTQDESQGTGDKFMGQGNNSLKINNFNTLEAQYIRLLFPVSCH 255

Qy 2177 IRSTLRMELMGCDLNSCMPLGMSKATSDAQITASSYFT--NMFA-TWSPSKARLHLQG 2233
Db 256 RGCTLRFELLGCELHGCCBEPGLKNTIPDSQITASSYKTNLRAFQWYPHLGLDNQG 315

Qy 2234 RSNAPRPQVNPKEWLOVDFOKMKVTGTTQVKSLLTSMVTKPELSSSQDGHQWTLF 2293
Db 316 KINAWTAQNSAKAWLQVDLQKQVVTGLITQAGDFGHQIVASVKVAHSDGQWTVY 375

Qy 2294 FQNGKVPQGNQDSFTPVNSLDPLLTLYLRHPQSVWHQIALRMEVLGC 2345
Db 376 EQGSKVPQGNLDNNSKKNIFFKFMARYVRVLPLSWENRITLRJELLGC 427

RESULT 12
A36479
C;Superfamily: milk fat globule membrane protein - mouse
C;Species: Mus musculus (house mouse)
C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
C;Accession: A36479
R;Stubbs, J.D.; Lekutis, C.; Singer, K.L.; Bui, A.; Yuzuki, D.; Sriniwasan, U.; Parry, G.
Proc. Natl. Acad. Sci. U.S.A. 87, 8417-8421, 1990
A;Title: cDNA cloning of a mouse mammary epithelial cell surface protein reveals the ex
A;Reference number: A36479; MUID:91046006; PMID:2122462
A;Accession: A36479
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-463 <STU>
A;Cross-references: GB:M38337; NID:G199142; PID:AAA39534.1; PID:G199143
C;Superfamily: milk fat globule protein; discoidin I amino-terminal homology; EGF homol
C;Keywords: membrane protein
F;68-107/Domain: EGF homology <EG1>
F;147-303/Domain: discoidin I amino-terminal homology <DN1>
F;307-463/Domain: discoidin I amino-terminal homology <DN2>

Query Match 5.3%; Score 657; DB 1; Length 463;
Best Local Similarity 42.2%; Pred. No. 6.9e-27;
Matches 135; Conservative 63; Mismatches 110; Indels 12; Gaps 5;

Qy 2037 SNKCTPLGASGHIRDFQITASQY-----GOWAPKLARLHYSGSINAW--STK2PFS 2088
Db 145 ASRCSTQLGMEGGAIAQSDISASVYVGMGLQWGPPELARLYRTGIVNAWHAGSNYDCKP 204

Qy 2089 WIKVDLLAPMIHGIKTQAGKQKTSLSIYISQIFIMYSLDGKKWQYRGNSTGLTMVFFGN 2148
Db 205 WIQVNLRLKRVSGVMTCQASAGRAEFLKTFKVAISLDGKFFQIDESGD-KEFLGN 263

Qy 2149 VDSGQIKHNIENPPIIARYIRLHPHYSIRSTLRMELMGCDLNSCMPLGMSKATSDAQ 2208
Db 264 LDNNSLKVMNFTLEAQYIRLVPVSCHRGCTLRPELLGCELHGCCLEPLGLKNTIPDSQ 323

Qy 2209 ITASSYFT--NMFA-TWSPSKARLHLOGRSNAPVQVNPKEWLOVDFOKMKVTGTTQ 2265
Db 324 MSASSYKTNLRAFQWYPHLGLDNQKINAWTAQNSAKEWLQVDLQGTQRCQVTHITQ 383

Qy 2266 GVKSLLTSMVTKPELSSSQDGHQWTLFFQNGKVKVQFQGNQDSFTPVVNSLDPLLTRYL 2325
Db 384 GARDFGHIQVESYKVAHSDGQWTVYEEQSGSKVFGQNLNNSHKKNIFKFPFARYV 443

Qy 2326 RIHQSVWHQIALRMEVLGC 2345
Db 444 RVLPSWENRITLRJELLGC 463

RESULT 13
T11743
P;47 protein - pig (fragment)
C;Species: Sus scrofa domestica (domestic pig)
C;Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 08-Sep-2002
C;Accession: T11743
R;Ensslin, M.; Vogel, T.; Calvete, J.J.; Thole, H.H.; Schmidtke, J.; Matsuda, T.; Toepfe,
Biol. Reprod. 58, 1057-1064, 1998
A;Title: Molecular cloning and characterization of P47, a novel boar sperm-associated zo
A;Reference number: 217325; MUID:98206817; PMID:9546740
A;Accession: T11743
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: mRNA
A;Residues: 1-409 <ENS>
A;Cross-references: EMBL:Y11683; NID:92652927; PIDN:CAA72379.1; PID:92652928
A;Experimental source: testis
C;Function:
A;Description: may be involved in membrane remodeling and/or function as a zona pellucid
C;Superfamily: milk fat globule protein; discoidin I amino-terminal homology; EGF homolo
F;6-40/Domain: EGF homology <EGF>

Query Match 5.2%; Score 650; DB 2; Length 409;
Best Local Similarity 39.3%; Pred. No. 1.3e-26;
Matches 139; Conservative 62; Mismatches 125; Indels 28; Gaps 6;

Qy 2018 ECLIGELHAGMSTLFLVYSNK-----COTPLGASGHIRDFQITASQY-- 2062
Db 58 ECEVIDDAHG--DVFTEYICKPHGYTGHCETICNAPLGMETGALADFOISASSMHG 115

Qy 2063 ----GOWAPKLARLHYSGSINAW--STKEPSWIKVDLLAPMIHGIKTQAGKFFSLY 2116
Db 116 FMGLQWAPELARLHRAGIENAWTASNYDRNPWQVNLRLRMRVGTGVTTQASRAGSAY 175

Qy 2117 ISQFLIMYSLDGKKQWYRGNSTGLMVFVFGVDSGGIKHNIENPPIIARYIRLHPHYS 2176
Db 176 MKTFKVAISTDGRKQFQIGAESGDKIFMGNLDNSGLKVNLFVPLEVQYVRLVPIICH 235

Qy 2177 IRSTLRMELMGCDLNSCMPLGMSKATSDAQITASSYFTN---MFATWSPSKARLHLQG 2233
Db 236 RGCTLRFELLGCELSCAPPLGLKNTIPNKQITASSFYRTNGLSFAFSWYFFYARDNQG 295

Qy 2234 RSNAPRPQVNPKEWLOVDFOKMKVTGTTQVKSLLTSMYKPELSSSQDGHQWTLF 2293
Db 296 KFNAAQNSASEMLOIDLGSQRRVTGIIITQARDFGHIQYVAAKVAISDDGVSTWEY 355

Qy 2294 FQNGKVPQGNQDSFTPVNSLDPLLTLYLRHPQSVWHQIALRMEVLGC 2345
Db 356 RDQALGKIFPQNLNNSHKKNMFETPLTRVRLPVAWHNRITLRVELLGC 409

RESULT 14

S65138
Glycoprotein antigen MG57/53, mammary gland - bovine (fragment)
N;Alternate names: Glycoprotein component 16/major fat-globule membrane protein/MFG-E8 h
C;Species: Bos primigenius taurus (cattle)
C;Date: 28-Oct-1996 #sequence_revision 13-Mar-1997 #text_change 07-Aug-1998
C;Accession: S65138; G48394
R;Aoki, N.; Kishi, Y.; Taniguchi, Y.; Adachi, T.; Nakamura, R.; Matsuda, T.
Biochim. Biophys. Acta 1245, 385-391, 1995

A:Title: Molecular cloning of glycoprotein antigens MGPS7/53 recognized by monoclonal anti
A:Reference number: S65138; MUID:96125736; PMID:8541316
A:Accession: S65138
A:Status: preliminary; not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 1-401 <NA>
R:Adachi, N.; Kishi, M.; Taniguchi, Y.; Adachi, T.; Nakamura, R.; Matsuda, T.
Biochim. Biophys. Acta 1245, 385-391, 1995
A:Title: Molecular cloning of glycoprotein antigens MGPS7/53 recognized by monoclonal anti
A:Reference number: S65138; MUID:96125736; PMID:8541316
A:Accession: S65138
A:Status: preliminary; not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 207-220 <NA>
R:Adachi, N.; Kishi, M.; Taniguchi, Y.; Adachi, T.; Nakamura, R.; Matsuda, T.
Biochim. Biophys. Acta 1245, 385-391, 1995
A:Title: The major fat-globule membrane proteins, bovine components 15/16 and guinea-pig
II-like sequences.
A:Reference number: A48394; MUID:93250576; PMID:8485470
A:Accession: G48394
A:Status: preliminary
A:Molecule type: protein
A:Experimental source: milk
A:Superfamily: milk fat globule protein; discoidin I amino-terminal homology; EGF homolo
C:Keywords: glycoprotein
F:1-32/Domain: EGF homology (fragment) <EG1>
F:40-79/Domain: EGF homology <EG2>
F:82-239/Domain: discoidin I amino-terminal homology <DN1>
F:243-401/Domain: discoidin I amino-terminal homology <DN2>
Query Match 5.1%; Score 635; DB 2; Length 401;
Best Local Similarity 37.9%; Pred. No. 7.8e-26;
Matches 135; Conservative 66; Mismatches 123; Indels 32; Gaps 7;
QY 2018 ECLIGHLHAGSTLFLVYSNK-----CQPLGASGHIRPQTASQY-- 2062
DB 50 ECQVTDSDHRG--DVPIQICPLGVGHCETCTSPGLMOTGAIADSIQASSMHLG 107
QY 2063 ----GOWAPKLARLHYSGSINAWST---KEPFSWIKVDLLAPMIHGIKTQCARQKFS 2114
DB 108 FMGLQWAPLAEALHGTGIVNAWTSNGYDKNP--WQVNLMRKMWVTGVVTCASRAGSA 165
QY 2115 LYISQFIIMYSLDGKKWQYRGNSGTGLMVFQGVNDSSGKIKNIPNPPIIARVIRLPH 2174
DB 166 EYLKTFKVAYSTDGRQFQFQVAGRSQDKIFIGNVNSGLKINLFDTELETQVRLVP 225
QY 2175 YSTRILRMELMGCDLNSCMPLGMSKASDAQITASSYFTN--MFATWSFSKARLHL 2231
DB 226 CHRGTCTLRFELGCELNGCTPELGLKNDITPNKQITASSYKWTGLSFAFSNFPYARLD 285
QY 2232 QGRSNARPOVNNPKWQLQVDFQKTMKVTGVTGQVKSLTSMYKVEFLISSQDGHQWT 2291
DB 286 QGRFNWATQNSASEWLQIDLSQKRVGTIIQTGARDFGHIQVVAAYRVAAGDDGVV 345
QY 2292 LFTQNG--KVKVFGQNDSTFPVNSLDPELLTRYLRIHPQSVVHQTALRMEVLGC 2345
DB 346 EYKDPGASESKIFFGNNDNNSHKKNIFETPFQARFVRIQVAVHNRIILRVLLG 401
RESULT 15
PAS-6/7 protein precursor - bovine
N:Alternate names: glycoprotein component 16/major fat-globule membrane protein/MFG-E8 h
C:Species: Bos primigenius taurus (cattle)
C:Date: 04-Dec-1997 #sequence revision 12-Dec-1997 #text_change 04-Nov-2002
C:Accession: S74211; S78114; S24181; S65138; G48394
R:Hvarregaard, J.; Andersen, M.H.; Berglund, L.; Rasmussen, J.T.; Petersen, T.E.
Eur. J. Biochem. 240, 628-636, 1996
A:Title: Characterization of glycoprotein PAS-6/7 from membranes of bovine milk fat glob
A:Reference number: S74211; MUID:97008954; PMID:8856064
A:Accession: S74211
A:Molecule type: mRNA
A:Residues: 1-427 <HVA>
A:Cross-references: EM3L:X91895; NID:gi632778; PIDN:CAA62997.1; PID:gi632779
A:Accession: S78114
A:Molecule type: protein
A:Residues: 19-85;96-110;140-165;174-216;221-232;248-277;285-293;309-337;339-420;425-427
R:Xim, D.H.; Kanno, C.; Mizokami, Y.
Biochim. Biophys. Acta 1122, 203-211, 1992

A:Title: Purification and characterization of major glycoproteins, PAS-6 and PAS-7, from
A:Reference number: S23926; MUID:92353107; PMID:1643094
A:Accession: S24181
A:Molecule type: protein
A:Residues: 383-394 <KIM>
R:Aoki, N.; Kishi, M.; Taniguchi, Y.; Adachi, T.; Nakamura, R.; Matsuda, T.
Biochim. Biophys. Acta 1245, 385-391, 1995
A:Title: Molecular cloning of glycoprotein antigens MGPS7/53 recognized by monoclonal anti
A:Reference number: S65138; MUID:96125736; PMID:8541316
A:Accession: S65138
A:Status: preliminary; not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 27-427 <AOK>
R:Mather, I.H.; Banghart, L.R.; Lane, W.S.
Biochem. Mol. Biol. Int. 29, 545-554, 1993
A:Title: The major fat-globule membrane proteins, bovine components 15/16 and guinea-pig
II-like sequences.
A:Reference number: A48394; MUID:93250576; PMID:8485470
A:Accession: G48394
A:Status: preliminary
A:Molecule type: protein
A:Residues: 233-246 <MAT>
A:Experimental source: milk
A:Note: sequence extracted from NCBI backbone (NCBI:P131457)
C:Superfamily: milk fat globule protein; discoidin I amino-terminal homology; EGF homolo
C:Keywords: blocked amino end; disulfide bond; glycoprotein; milk
F:1-18/Domain: signal sequence #status predicted <SIG>
F:19-427/Product: PAS-6/7 protein #status experimental <MAT>
F:24-38/Domain: EGF homology <EG1>
F:56-105/Domain: EGF homology <EG2>
F:108-265/Domain: discoidin I amino-terminal homology <DN1>
F:269-427/Domain: discoidin I amino-terminal homology <DN2>
F:24-35,29-47,49-58,66-77,71-94,96-105/Disulfide bonds: #status predicted
F:27/Binding site: carbohydrate (Ser) (covalent) #status experimental
F:34/Binding site: carbohydrate (Thr) (covalent) #status experimental
F:59,227/Binding site: carbohydrate (Asn) (covalent) #status experimental
F:109-265,252-256,270-427/Disulfide bonds: #status experimental
Query Match 5.1%; Score 635; DB 2; Length 427;
Best Local Similarity 37.9%; Pred. No. 8.6e-26;
Matches 135; Conservative 66; Mismatches 123; Indels 32; Gaps 7;
QY 2018 ECLIGHLHAGSTLFLVYSNK-----CQPLGASGHIRPQTASQY-- 2062
DB 76 ECQVTDSDHRG--DVPIQICPLGVGHCETCTSPGLMOTGAIADSIQASSMHLG 133
QY 2063 ----GOWAPKLARLHYSGSINAWST---KEPFSWIKVDLLAPMIHGIKTQCARQKFS 2114
DB 134 FMGLQWAPLAEALHGTGIVNAWTSNGYDKNP--WQVNLMRKMWVTGVVTCASRAGSA 191
QY 2115 LYISQFIIMYSLDGKKWQYRGNSGTGLMVFQGVNDSSGKIKNIPNPPIIARVIRLPH 2174
DB 192 EYLKTFKVAYSTDGRQFQFQVAGRSQDKIFIGNVNSGLKINLFDTELETQVRLVP 251
QY 2175 YSTRILRMELMGCDLNSCMPLGMSKASDAQITASSYFTN--MFATWSFSKARLHL 2231
DB 252 CHRGTCTLRFELGCELNGCTPELGLKNDITPNKQITASSYKWTGLSFAFSNFPYARLD 311
QY 2232 QGRSNARPOVNNPKWQLQVDFQKTMKVTGVTGQVKSLTSMYKVEFLISSQDGHQWT 2291
DB 312 QGRFNWATQNSASEWLQIDLSQKRVGTIIQTGARDFGHIQVVAAYRVAAGDDGVV 371
QY 2292 LFTQNG--KVKVFGQNDSTFPVNSLDPELLTRYLRIHPQSVVHQTALRMEVLGC 2345
DB 372 EYKDPGASESKIFFGNNDNNSHKKNIFETPFQARFVRIQVAVHNRIILRVLLG 427
Search completed: April 13, 2004, 14:09:55
Job time : 44 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 13, 2004, 14:01:27 ; Search time 18 seconds
(without alignments)
6800.941 Million cell updates/sec

Title: NP000123-328-355-581

Perfect score: 12415

Sequence: 1 YQIELSTCFPLCLLRPFSA.....WVHQIALRMEVLGCEAQDLY 2351

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_42.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	12399	99.9	2351	1 FAS_HUMAN	P00451 homo sapien
2	8820	71.0	2319	1 FAS_MOUSE	Q06194 mus musculus
3	8057	64.9	2133	1 FAS_PIG	P12263 sus scrofa
4	2823	5	2224	1 FAS_HUMAN	P12259 homo sapien
5	2776	5	2258	1 FAS_PIG	Q28107 sus scrofa
6	2743	22.1	2211	1 FAS_BOVIN	Q28107 bos taurus
7	1322	10.6	1065	1 CERU_HUMAN	P00450 homo sapien
8	1274	10.3	1059	1 CERU_MOUSE	P13635 rattus norv
9	1243	10.0	1062	1 CERU_MOUSE	Q61147 mus musculus
10	663	5.3	427	1 MFGM_MOUSE	P70490 rattus norv
11	657	5.3	463	1 MFGM_MOUSE	P21956 mus musculus
12	650	5.2	409	1 MFGM_PIG	P79385 sus scrofa
13	635	5.1	427	1 MFGM_BOVIN	Q95114 bos taurus
14	588	4.7	387	1 MFGM_HUMAN	Q08431 homo sapien
15	469	3.8	931	1 NRP2_HUMAN	O60462 homo sapien
16	464	3.7	925	1 NRP2_MOUSE	O35276 rattus norv
17	462	3.7	931	1 NRP2_MOUSE	O35375 mus musculus
18	458	3.7	924	1 NRP1_MOUSE	P79795 gallus gall
19	451	3.6	922	1 NRP1_MOUSE	Q9qwj9 rattus norv
20	446	3.6	923	1 NRP1_MOUSE	P77333 mus musculus
21	443	3.6	928	1 NRP1_XENLA	P28824 xenopus lae
22	429	3.5	923	1 NRP1_HUMAN	O14786 homo sapien
23	306	2.5	3133	1 HMCT_BOMMO	P98092 bombyx mori
24	266	2.1	764	1 CPX2_MOUSE	Q9d215 mus musculus
25	261	2.1	756	1 CPX2_HUMAN	Q8n436 homo sapien
26	260	2.1	280	1 XLR1_FUGRU	Q9w6r5 fugu rubrip
27	247	2.0	224	1 XLR1_MOUSE	Q9z114 mus musculus
28	243	2.0	224	1 XLR1_HUMAN	O15537 homo sapien
29	242	2.0	3418	1 SRC2_HUMAN	P51587 homo sapien
30	235	1.9	1358	1 SIR4_YEAST	P11978 saccharomyc
31	223	1.8	722	1 CPXM_MOUSE	Q9z100 mus musculus
32	221	1.8	3329	1 BRC2_MOUSE	P97929 mus musculus
33	219	1.8	1928	1 MYS1_YEAST	P08964 saccharomyc

RESULT 1				
FAS_HUMAN				
ID	FAS_HUMAN	STANDARD;	PRT;	2351 AA.
AC	P00451;			
DT	21-JUL-1986	(Rel. 01, Created)		
DT	21-JUL-1986	(Rel. 01, Last sequence update)		
DT	10-OCT-2003	(Rel. 42, Last annotation update)		
DE	Coagulation factor VIII precursor (Procoagulant component)			
DE	(Antihemophilic factor) (AHF).			
GN	F8 OR F8C.			
OS	Homo sapiens (Human).			
OS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=86081164; PubMed=3935400;			
RA	Truett M.A., Blacher R., Burke R.L., Caput D., Chu C., Dina D.,			
RA	Hartog K., Kuo C.H., Masiaz F.R., Merryweather J.P., Najarian R.,			
RA	Pachl C., Potter S.J., Puma J., Quiroga M., Rall L.B., Randolph A.,			
RA	Urdea M.S., Valenzuela P., Dahl H.-H.M., Favalaro J., Hansen J.,			
RA	Nordfang O., Ezban M.;			
RT	"Characterization of the polypeptide composition of human factor VIII:C and the nucleotide sequence and expression of the human kidney cDNA."			
RT	DNA 4:333-349 (1985).			
RL	[2]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=85061548; PubMed=6438526;			
RA	Wood W.I., Capon D.J., Simonsen C.C., Eaton D.L., Gitschier J.,			
RA	Keyt B., Seeburg P.H., Smith D.H., Hollingshead P., Wion K.L.,			
RA	Delwart E., Tuddenham E.G.D., Vehar G.A., Lawn R.M.;			
RT	"Expression of active human factor VIII from recombinant DNA clones."			
RL	Nature 312:330-337 (1984).			
RL	[3]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=85061550; PubMed=6438528;			
RA	Toole J.J., Knopf J.L., Wozney J.M., Sultzman L.A., Buecker J.L.,			
RA	Pittman D.D., Kaufman R.J., Brown E., Shoemaker C., Orr E.C.,			
RA	Amphlett G.W., Foster W.B., Coe M.L., Knutson G.J., Pass D.N.,			
RA	Hewick R.M.;			
RT	"Molecular cloning of a cDNA encoding human antihemophilic factor."			
RL	Nature 312:342-347 (1984).			
RL	[4]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=93265012; PubMed=1303178;			
RA	Gitschier J., Wood W.I.;			
RT	"Sequence of the exon-containing regions of the human factor VIII gene."			
RL	Hum. Mol. Genet. 1:199-200 (1992).			
RL	[5]			
RP	SEQUENCE OF 2064-2070 FROM N.A.			
RA	de Water N.S., Williams R., Browett P.J.;			
RL	submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.			
RL	[6]			
RP	SULFATION OF TYR-1699.			

P47580 mycoplasma
Q96sm3 homo sapien
Q9ut06 homo sapien
P34216 saccharomyc
Q01484 homo sapien
Q01613 xenopus lae
Q10411 schistosacch
Q94887 drosophila
Q00799 plasmodium
P47549 mycoplasma
Q99p47 mus musculu
P20929 homo sapien

ALIGNMENTS

RX MEDLINE=91093266; PubMed=1898735;
RA Leyte A.; van Schijndel H.B.; Niehrs C., Huttner W.B., Verbeet M.P.,
RA Metters K., van Mourik J.A.;
RT "Sulfation of Tyr1680 of human blood coagulation factor VIII is
RT essential for the interaction of factor VIII with von Willebrand
RT factor.";
RL J. Biol. Chem. 266:740-746(1991).
RN [7]
RP SULFATION.
RX MEDLINE=92207952; PubMed=154716;
RA Pittman D.D., Wang J.H., Kaufman R.J.;
RA "Identification and functional importance of tyrosine sulfate
RT residues within recombinant factor VIII.";
RL Biochemistry 31:3315-3325(1992).
RN [8]
RP STRUCTURE BY NMR OF 2322-2343.
RX MEDLINE=95200924; PubMed=7893714;
RA Gilbert G.E., Baleja J.D.;
RA "Membrane-binding peptide from the C2 domain of factor VIII forms an
RT amphipathic structure as determined by NMR spectroscopy.";
RL Biochemistry 34:3022-3031(1995).
RN [9]
RP REVIEW ON MOLECULAR BASIS OF HEMA.
RX MEDLINE=91221499; PubMed=1902642;
RA Gitschier J.;
RA "The molecular basis of hemophilia A.";
RL Ann. N.Y. Acad. Sci. 614:89-96(1991).
RN [10]
RP REVIEW ON MOLECULAR BASIS OF HEMA.
RX MEDLINE=89088506; PubMed=2491949;
RA White G.C. II, Shoemaker C.B.;
RA "Factor VIII gene and hemophilia A.";
RL Blood 73:1-12(1989).
RN [11]
RP REVIEW ON MOLECULAR BASIS OF HEMA.
RX MEDLINE=95245332; PubMed=7728145;
RA Antonarakis S.E., Kazazian H.H., Tuddenham E.G.D.;
RT "Molecular etiology of factor VIII deficiency in hemophilia A.";
RL Hum. Mutat. 5:1-22(1995).
RN [12]
RP VARIANT HEMA GLN-2326.
RX MEDLINE=86235434; PubMed=3012775;
RA Gitschier J., Wood W.I., Shuman M.A., Lawn R.M.;
RA "Identification of a missense mutation in the factor VIII gene of a
RT mild hemophilic.";
RL Science 232:1415-1416(1986).
RN [13]
RP VARIANT HEMA PRO-2135.
RX MEDLINE=88096539; PubMed=3122181;
RA Levinson B., Janco R.L., Phillips J.A. III, Gitschier J.;
RA "A novel missense mutation in the factor VIII gene identified by
RT analysis of amplified hemophilia DNA sequences.";
RL Nucleic Acids Res. 15:9797-9805(1987).
RN [14]
RP VARIANT HEMA GLN-2228.
RX MEDLINE=86191889; PubMed=2833855;
RA Yousoufian H., Antonarakis S.E., Bell W., Griffin A.M.,
RA Kazazian H.H.;
RT "Nonsense and missense mutations in hemophilia A: estimate of the
RT relative mutation rate at CG dinucleotides.";
RL Am. J. Hum. Genet. 42:718-725(1988).
RN [15]
RP VARIANT HEMA GLY-291.
RX MEDLINE=89220354; PubMed=2835904;
RA Yousoufian H., Wong C., Aronis S., Platakoukis H., Kazazian H.H. Jr.,
RA Antonarakis S.E.;
RT "Moderately severe hemophilia A resulting from Glu-->Gly substitution
RT in exon 7 of the factor VIII gene.";
RL Am. J. Hum. Genet. 42:867-871(1988).
RN [16]
RP VARIANT HEMA CYS-1708.
RX MEDLINE=89274393; PubMed=2499363;
RA C'Brien D.P., Tuddenham E.G.;
RT "Purification and characterization of factor VIII 1,689-Cys: a
RT nonfunctional cofactor occurring in a patient with severe hemophilia
RL A.";
RL Blood 73:2117-2122(1989).
RN [17]
RP VARIANT HEMA CYS-391.
RX MEDLINE=90001543; PubMed=2506948;
RA Shima M., Ware J., Yoshioka A., Fukui H., Pulcher C.A.;
RA "An arginine to cysteine amino acid substitution at a critical
RT thrombin cleavage site in a dysfunctional factor VIII molecule.";
RL Blood 74:1612-1617(1989).
RN [18]
RP VARIANT HEMA LEU-189.
RX MEDLINE=90057680; PubMed=2510835;
RA Chan V., Chan T.K., Tong T.M., Todd D.;
RA "A novel missense mutation in exon 4 of the factor VIII:C gene
RT resulting in moderately severe hemophilia A.";
RL Blood 74:2688-2691(1989).
RN [19]
RP VARIANT HEMA LEU-2326.
RX MEDLINE=89197216; PubMed=2495245;
RA Inaba H., Fujimaki M., Kazazian H.H. Jr., Antonarakis S.E.;
RA "Mild hemophilia A resulting from Arg-to-Leu substitution in exon 26
RT of the factor VIII gene.";
RL Hum. Genet. 81:335-338(1989).
RN [20]
RP VARIANT HEMA HIS-391.
RX MEDLINE=89264602; PubMed=2498882;
RA Arai M., Inaba H., Higuchi M., Antonarakis S.E., Kazazian H.H. Jr.,
RA Fujimaki M., Hoyer L.W.;
RT "Direct characterization of factor VIII in plasma: detection of a
RT mutation altering a thrombin cleavage site
RT (arginine-372-->histidine).";
RL Proc. Natl. Acad. Sci. U.S.A. 86:4277-4281(1989).
RN [21]
RP VARIANT HEMA CYS-1708.
RX MEDLINE=90105723; PubMed=2104766;
RA Arai M., Higuchi M., Antonarakis S.E., Kazazian H.H. Jr.,
RA Phillips J.A. III, Janco R.L., Hoyer L.W.;
RT "Characterization of a thrombin cleavage site mutation (Arg 1689 to
RT Cys) in the factor VIII gene of two unrelated patients with
RT cross-reacting material-positive hemophilia A.";
RL Blood 75:384-389(1990).
RN [22]
RP VARIANTS HEMA GLN-2228 AND LEU-2326.
RX MEDLINE=90123183; PubMed=2105106;
RA Casula L., Murru S., Pecorara M., Ristaldi M.S., Restagno G.,
RA Mancuso G., Viorini M., de Biasi R., Baudo F., Carbonara A.;
RT "Recurrent mutations and three novel rearrangements in the factor
RT VIII gene of hemophilia A patients of Italian descent.";
RL Blood 75:662-670(1990).
RN [23]
RP VARIANT HEMA CYS-391.
RX MEDLINE=90329422; PubMed=1973901;
RA Pattinson J.K., McVey J.H., Boon M., Ajani A., Tuddenham E.G.;
RT "CRM+ hemophilia A due to a missense mutation (372-->Cys) at the
RT internal heavy chain thrombin cleavage site.";
RL Br. J. Haematol. 75:73-77(1990).
RN [24]
RP VARIANTS HEMA PHE-1699 AND CYS-1708.
RX MEDLINE=90152691; PubMed=2105906;
RA Higuchi M., Wong C., Kochan L., Olek K., Aronis S., Kasper C.K.,
RA Kazazian H.H., Antonarakis S.E.;
RT "Characterization of mutations in the factor VIII gene by direct
RT sequencing of amplified genomic DNA.";
RL Genomics 6:65-71(1990).
RN [25]
RP VARIANTS HEMA CYS-1728 AND ASP-1941.
RX MEDLINE=90169985; PubMed=2106480;
RA Traystman M.D., Higuchi M., Kasper C.K., Antonarakis S.E.,
RA Kazazian H.H.;
RT "Use of denaturing gradient gel electrophoresis to detect point
RT mutations in the factor VIII gene.";

Query Match 99.9%; Score 12399; DB 1; Length 2351;									
Best Local Similarity 99.9%; Pred. NO. 0;									
Matches 2348; Conservative 1; Mismatches 2; Indels 0; Gaps 0;									
QY	1	MQIELSTCFCLLRFCFSATRYVYGAVELSDVMQSDLGELPVDARTPPVPKSPFPN	60						
DB	1	MQIELSTCFCLLRFCFSATRYVYGAVELSDVMQSDLGELPVDARTPPVPKSPFPN	60						
QY	61	TSVVYKKTFLVEFTDHLFNAIKPRPPWMLGLGPTIQAEVYDVTVVITLKNMASHPVLSHAV	120						
DB	61	TSVVYKKTFLVEFTDHLFNAIKPRPPWMLGLGPTIQAEVYDVTVVITLKNMASHPVLSHAV	120						
QY	121	GVSYKASBGAEEYDDQTSREKDDKVFPQGSHTYVQVLKENGPKASDPLCLITYLSLH	180						
DB	121	GVSYKASBGAEEYDDQTSREKDDKVFPQGSHTYVQVLKENGPKASDPLCLITYLSLH	180						
QY	181	VDLVKDLNSGLIGALLVCREGSLAKETQTLHKFILLFAVDEGKSWHSETKNSLMQDRD	240						
DB	181	VDLVKDLNSGLIGALLVCREGSLAKETQTLHKFILLFAVDEGKSWHSETKNSLMQDRD	240						
QY	241	AASARAMPKHTVNGYVNRSLPGLIGCHRXSVYVHVIGMGITPEVESIFLEGHTFLVRNH	300						
DB	241	AASARAMPKHTVNGYVNRSLPGLIGCHRXSVYVHVIGMGITPEVESIFLEGHTFLVRNH	300						
QY	301	QOASLEISPTITLTAQTLMDLGGFLISCHISSHQHGMAYVVDSCPEPQURMKNE	360						
DB	301	QOASLEISPTITLTAQTLMDLGGFLISCHISSHQHGMAYVVDSCPEPQURMKNE	360						
QY	361	EADYDDDLTDEMDVRFDDNPSFTQIRSAKHKPKTWVHYIAAEEDDWDYAPLVLA	420						
DB	361	EADYDDDLTDEMDVRFDDNPSFTQIRSAKHKPKTWVHYIAAEEDDWDYAPLVLA	420						
QY	421	PDDRSYKQYLNNGPORIGRYKXKVRNAYTDTFTKREAIQHESGILGPLLYGEVGDTL	480						
DB	421	PDDRSYKQYLNNGPORIGRYKXKVRNAYTDTFTKREAIQHESGILGPLLYGEVGDTL	480						
QY	481	LIIFPNQASRYNYPHGITDVRPLYSRPLPKGVKHLKDFPILCEIPFKYKWTVTVEDGP	540						
DB	481	LIIFPNQASRYNYPHGITDVRPLYSRPLPKGVKHLKDFPILCEIPFKYKWTVTVEDGP	540						
QY	541	TKSDPRCLTRYSSVFNMERDLASGLIGPLIICYKESVDQGNQIMSKRVILFSVFDE	600						
DB	541	TKSDPRCLTRYSSVFNMERDLASGLIGPLIICYKESVDQGNQIMSKRVILFSVFDE	600						
QY	601	NRSWYLTENIORFLPNPAGVQLEDPEFOASNIMHSINGVFDLSQLSVCLHEVAYWYILS	660						
DB	601	NRSWYLTENIORFLPNPAGVQLEDPEFOASNIMHSINGVFDLSQLSVCLHEVAYWYILS	660						
QY	661	IGAOTDELVSFPSSGYTFKHKVYEDTLTLPFSGTTFVMSNPNGLWILGCHNSDFNRG	720						
DB	661	IGAOTDELVSFPSSGYTFKHKVYEDTLTLPFSGTTFVMSNPNGLWILGCHNSDFNRG	720						
QY	721	MTALLKVSCKNTGDYEDSYEDISAYLLSKNNAIEPSPQNSRHPSTROKQFNATTI	780						
DB	721	MTALLKVSCKNTGDYEDSYEDISAYLLSKNNAIEPSPQNSRHPSTROKQFNATTI	780						
QY	781	PENDIEKTDWFAHRTMPKIQNVSSDILLMLLRQSPTHGLSLDQAKYETESDDPS	840						
DB	781	PENDIEKTDWFAHRTMPKIQNVSSDILLMLLRQSPTHGLSLDQAKYETESDDPS	840						
QY	841	PGAIIDNNNSISEMTHFRPOLHSGDMVFTPEGLQLRLNEKLGTTAATLKKLDFKVSST	900						
DB	841	PGAIIDNNNSISEMTHFRPOLHSGDMVFTPEGLQLRLNEKLGTTAATLKKLDFKVSST	900						
QY	901	SNKLISTIPSDNLAAGTNTSSIGPPSPVHYDQSDTLTGKKSPLTESGGLSLSEB	960						
DB	901	SNKLISTIPSDNLAAGTNTSSIGPPSPVHYDQSDTLTGKKSPLTESGGLSLSEB	960						
QY	961	NNDKXILLESGLMNSQSSWGKNSVSTESGELFKGKAHGPALLTKONALFKVSIISLKTN	1020						
DB	961	NNDKXILLESGLMNSQSSWGKNSVSTESGELFKGKAHGPALLTKONALFKVSIISLKTN	1020						
QY	1021	KTSNNSATNRKTHIDGPSLLIENSFSVWQNLILESDETEFKKVTPLIHDRMLMDKNAIALRL	1080						
DB	1021	KTSNNSATNRKTHIDGPSLLIENSFSVWQNLILESDETEFKKVTPLIHDRMLMDKNAIALRL	1080						
QY	1081	NKMSNKTSSKXNEMVQOKKEGPIPPDAQNPMSFPKMLFLPESARWIORTHGKNSLNSG	1140						
DB	1081	NKMSNKTSSKXNEMVQOKKEGPIPPDAQNPMSFPKMLFLPESARWIORTHGKNSLNSG	1140						
QY	1141	QGPSKQLVSLGPESXVBQGNFLSEKXKVVVGKGEFTKDVGLKEMVFPSSRNJFTNLDN	1200						
DB	1141	QGPSKQLVSLGPESXVBQGNFLSEKXKVVVGKGEFTKDVGLKEMVFPSSRNJFTNLDN	1200						
QY	1201	LEENHTNQEKKIOEBIEKKFWLIIQENVVLPOIHVTCTKNPMKMLFLLSTRQNVGSGYD	1260						
DB	1201	LEENHTNQEKKIOEBIEKKFWLIIQENVVLPOIHVTCTKNPMKMLFLLSTRQNVGSGYD	1260						
QY	1261	GAYAPVLOPFRSLNDSTNRTKHTAHFSKKGBEENLEGLGNOTKOIVEKYACTRISPNT	1320						
DB	1261	GAYAPVLOPFRSLNDSTNRTKHTAHFSKKGBEENLEGLGNOTKOIVEKYACTRISPNT	1320						
QY	1321	SOQNFVTOQSKRALKQFRPLLEBTELEKRIIVDDTSTQMSKMKHILTPSLTQIDYNEKE	1380						
DB	1321	SOQNFVTOQSKRALKQFRPLLEBTELEKRIIVDDTSTQMSKMKHILTPSLTQIDYNEKE	1380						
QY	1381	KGAIQSPJSDCLUTRSHSIPQANRSPILPAKVSFPSSIRPIVLTBVLFDONSSHLPAASY	1440						
DB	1381	KGAIQSPJSDCLUTRSHSIPQANRSPILPAKVSFPSSIRPIVLTBVLFDONSSHLPAASY	1440						
QY	1441	RKXDSGVCESSHFLQGAKKNNLSAILITLMTGDQREVGLSTSATNSVTYKVENVLP	1500						
DB	1441	RKXDSGVCESSHFLQGAKKNNLSAILITLMTGDQREVGLSTSATNSVTYKVENVLP	1500						
QY	1501	KPDLPKTSGKVELLPKVHIYQKOLPPTETNSGPGHLDLVEGSLLOQTEGAIKWNEANRP	1560						
DB	1501	KPDLPKTSGKVELLPKVHIYQKOLPPTETNSGPGHLDLVEGSLLOQTEGAIKWNEANRP	1560						
QY	1561	GKVPFLVATESSAKTSPKLLDPLANDNHYGTQIPKEWKSQESPKTAKKXDTILSL	1620						
DB	1561	GKVPFLVATESSAKTSPKLLDPLANDNHYGTQIPKEWKSQESPKTAKKXDTILSL	1620						
QY	1621	NACESNIAIAINNGONKPEIEVTWAKQGRTERLCSQNPVVLKSHOREITRTTLOSQOE	1680						
DB	1621	NACESNIAIAINNGONKPEIEVTWAKQGRTERLCSQNPVVLKSHOREITRTTLOSQOE	1680						
QY	1681	IDYDDTISVEMKKEDFDIYDEDENSQSPRSFOKTRHYFIAAVERLWDYGMSSSPEVLNRP	1740						
DB	1681	IDYDDTISVEMKKEDFDIYDEDENSQSPRSFOKTRHYFIAAVERLWDYGMSSSPEVLNRP	1740						
QY	1741	AQSGSVPOFKVVFQEFDDGSFTQPLXRGELNEHLGILGPVIRAEVEDNIWTFRNQASR	1800						
DB	1741	AQSGSVPOFKVVFQEFDDGSFTQPLXRGELNEHLGILGPVIRAEVEDNIWTFRNQASR	1800						
QY	1801	PYSFYSSLI SYEEDORQGAEPKRNKFNENKTYFWKYQHMAPTKDRPDCKAWAYFSDV	1860						
DB	1801	PYSFYSSLI SYEEDORQGAEPKRNKFNENKTYFWKYQHMAPTKDRPDCKAWAYFSDV	1860						
QY	1861	DLEKDVHSLGLPLLVCHTNTLPAHGQVTVQEPALPTTFDETKSWYFTENMERNORA	1920						
DB	1861	DLEKDVHSLGLPLLVCHTNTLPAHGQVTVQEPALPTTFDETKSWYFTENMERNORA	1920						
QY	1921	PCNIQWEDPTFKENYRFAHNGYIMDTLPLGMAQDQIRWVLLSNGSNENIHSIFSGH	1980						
DB	1921	PCNIQWEDPTFKENYRFAHNGYIMDTLPLGMAQDQIRWVLLSNGSNENIHSIFSGH	1980						
QY	1981	VFTVAKKBEYKVALXNLYPGVFEVEMLPKAGIWRVECLIGEHLHAGMSTFLVYSNKC	2040						
DB	1981	VFTVAKKBEYKVALXNLYPGVFEVEMLPKAGIWRVECLIGEHLHAGMSTFLVYSNKC	2040						
QY	2041	QTPFLGASGHIDPQITAGSGQGWAPKLARLHYGSSINAKSTKBPFSKIKYDLLAPMLI	2100						
DB	2041	QTPFLGASGHIDPQITAGSGQGWAPKLARLHYGSSINAKSTKBPFSKIKYDLLAPMLI	2100						
QY	2101	HGKITQGAQKQSSSLYSIQFIIMYSLDGGKWKQTYRGNSGTGLMVFVGNVDSSGIKHNTFN	2160						
DB	2101	HGKITQGAQKQSSSLYSIQFIIMYSLDGGKWKQTYRGNSGTGLMVFVGNVDSSGIKHNTFN	2160						

Wed Apr 14 15:52:00 2004

np000123-328-355-581.rsp

Db 2101 HGKTCARQKFSLSQISQIIMISLDGKKQYIRNGSTGLMVFNGVDSGQKHNFN 2160
 Qy 2161 PPIIARYIRLPHYSIRSLRMELMGCDLNSCMPJGMSKSAISDAQITASSYFTNMPA 2220
 Db 2161 PPIIARYIRLPHYSIRSLRMELMGCDLNSCMPJGMSKSAISDAQITASSYFTNMPA 2220
 Qy 2221 TWSPSKARLHQGRSNARQVNNPKWLQVDFQKIMKVTGVTTOGVKSLTSMYKKEFL 2280
 Db 2221 TWSPSKARLHQGRSNARQVNNPKWLQVDFQKIMKVTGVTTOGVKSLTSMYKKEFL 2280
 Qy 2281 ISSQDGHQWTLFFQNGKVKVQFQNGQDSFTPVNSLDPLLTRYLRIHPQSWVHQIALRM 2340
 Db 2281 ISSQDGHQWTLFFQNGKVKVQFQNGQDSFTPVNSLDPLLTRYLRIHPQSWVHQIALRM 2340
 Qy 2341 EVLGCRAQDLY 2351
 Db 2341 EVLGCRAQDLY 2351

RESULT 2
 ID _FAS MOUSE STANDARD; PRT; 2319 AA.
 AC Q06194;
 DT 01-JUN-1994 (Rel. 29, Created)
 DT 01-JUN-1994 (Rel. 29, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Coagulation factor VIII precursor (Procoagulant component).
 GN P8 OR CF8 OR P8C.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6 X CBA; TISSUE=Liver;
 RX MEDLINE=93300511; PubMed=8314577;
 RA Elder S., Lakich D., Gitschier J.;
 RT "Sequence of the murine factor VIII cDNA";
 RL Genomics 16:374-379(1993).
 CC -!- FUNCTION: Factor VIII, along with calcium and phospholipid, acts
 CC as a cofactor for factor IXa when it converts factor X to the
 CC activated form, factor Xa.
 CC -!- SUBCELLULAR LOCATION: Extracellular.
 CC -!- TISSUE SPECIFICITY: Found in most tissues.
 CC -!- SIMILARITY: Contains 3 F5/8 type A domains.
 CC -!- SIMILARITY: Contains 2 F5/8 type C domains.
 CC -!- SIMILARITY: STRONG, TO COAGULATION FACTOR V.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>)
 CC or send an email to license@sb-sib.ch.
 CC -----
 CC EMBL; L05573; AAA37385.1; --
 CC PIR; A47004; A47004.
 CC HSP; P00451; 1CFG.
 CC MGD; MGI:88383; F8.
 CC InterPro; IPR001117; Cu-oxidase.
 CC InterPro; IPR008972; Cupredoxin.
 CC InterPro; IPR000421; F5/8 C.
 CC InterPro; IPR008979; Gal_bind like.
 CC Pfam; P00394; Cu-oxidase; 3.
 CC Pfam; PF00754; F5 F8 type C; 2.
 CC SMART; SM00231; F5/8C_2.
 CC PROSITE; PS01285; F5/8C_1; 2.
 CC PROSITE; PS01286; F5/8C_2; 2.
 CC PROSITE; PS00022; F5/8C_3; 2.
 CC PROSITE; PS00079; MULTICOPPER_OXIDASE1; 3.
 CC Blood coagulation; Repeat; Plasma; Acute phase; Calcium;

KW Signal; Glycoprotein; Sulfation.
 FT SIGNAL 1 19 POTENTIAL.
 FT CHAIN 20 2319 COAGULATION FACTOR VIII.
 FT DOMAIN 20 349 F5/8 TYPE A 1.
 FT DOMAIN 20 349 PLASTOCYANIN-LIKE 1.
 FT DOMAIN 207 349 PLASTOCYANIN-LIKE 2.
 FT DOMAIN 399 730 F5/8 TYPE A 2.LIKE 3.
 FT DOMAIN 399 730 PLASTOCYANIN-LIKE 3.
 FT DOMAIN 583 730 PLASTOCYANIN-LIKE 4.
 FT DOMAIN 760 1640 B.
 FT DOMAIN 1683 2008 F5/8 TYPE A 3.
 FT DOMAIN 1683 1845 PLASTOCYANIN-LIKE 5.
 FT DOMAIN 1855 2008 PLASTOCYANIN-LIKE 6.
 FT DOMAIN 2008 2156 F5/8 TYPE C 1.
 FT DOMAIN 2161 2313 F5/8 TYPE C 2.
 FT SITE 391 392 CLEAVAGE (BY THROMBIN) (BY SIMILARITY).
 FT SITE 1678 1679 CLEAVAGE (BY THROMBIN) (BY SIMILARITY).
 FT SITE 1324 1325 CLEAVAGE (BY THROMBIN) (BY SIMILARITY).
 FT SITE 1640 1641 CLEAVAGE (ACTIVATION) (BY SIMILARITY).
 FT MOD_RES 367 367 SULFATION (BY SIMILARITY).
 FT MOD_RES 737 737 SULFATION (BY SIMILARITY).
 FT MOD_RES 738 738 SULFATION (BY SIMILARITY).
 FT MOD_RES 742 742 SULFATION (BY SIMILARITY).
 FT MOD_RES 1669 1669 SULFATION (REQUIRED FOR VWF BINDING) (BY SIMILARITY).
 FT MOD_RES 1687 1687 SULFATION (BY SIMILARITY).
 FT DISULFID 173 199 PROBABLE.
 FT DISULFID 547 573 PROBABLE.
 FT DISULFID 1819 1845 PROBABLE.
 FT DISULFID 2008 2156 BY SIMILARITY.
 FT DISULFID 2161 2313 BY SIMILARITY.
 FT CARBOHYD 61 61 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 233 233 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 259 259 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 423 423 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 601 601 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 880 880 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 958 958 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 1015 1015 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 1022 1022 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 1026 1026 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 1044 1044 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 1076 1076 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 1087 1087 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 1136 1136 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 1161 1161 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 1192 1192 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 1255 1255 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 1268 1268 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 1273 1273 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 1274 1274 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 1302 1302 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 1316 1316 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 1340 1340 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 1378 1378 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 1797 1797 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 2105 2105 N-LINKED (GLCNAC. .) (POTENTIAL).
 SQ SEQUENCE 2319 AA; 266148 MW; F0054DE051DB2A01 CRC64;
 Query Match 71.0%; Score 8820; DB 1; Length 2319;
 Best Local Similarity 72.4%; Pred. No. 0;
 Matches 1711; Conservative 232; Mismatches 365; Indels 54; Gaps 22;
 Qy 1 MQIELSTCFCLRLRFFCFSAIRRYTIGAVELSDWYMQSD-LGELPDVADPPRPVPSPPF 59
 Db 1 MQIALFACPFLLSFLNFCSSAIRRYLIGAVELSDWYMQSDLLSVLHDSRLPDMSTSPFP 60
 Qy 60 NTSVWYKTLFVEFTDLHFNIAKPRPWGCLLGTTCQAEVYDVTWTLKNWASHPSLSHA 119
 Db 61 NTSVWYKTLFVEFTDLHFNIAKPRPWGCLLGTTCQAEVYDVTWTLKNWASHPSLSHA 120
 Qy 120 VGSYWKASGEAEYDDQTSOREKDDKVPFGSGHTYVQVQLKENGPMASDFCLTYSYLS 179

Db 121 VGSYWKASDEYEDQISOMKEDDKVFPGESHTYVWVLEKNGPMBASDEPPCLTYSMS 180
Qy 180 HVDLVKOLNSGLIGALLVCREGLAKKKTQTLKXPIILLFAVDFGKGWHSBETQNSLMQDR 239
Db 181 HVDLVKOLNSGLIGALLVCREGLSKRTQMLYQVLLFAVDFGKGWHSBETQNSLMQDR 240
Qy 240 DAAGARAPKXHTVNGYVNRSLPGLICHRKSVYVHVIGMGTTPPEVASIFLEGHTELVRN 299
Db 241 DSASARDPKMETVNGYVNRSLPGLICHRKSVYVHVIGMGTTPPEVASIFLEGHTELVRN 300
Qy 300 HQASLETSPIPTTAOTLLMDLGOFLSCHISSHQHDMGMEAYVKVDCSPEPQIMK-N 358
Db 301 HQASLETSPIPTTAOTLLMDLGOFLSCHISSHQHDMGMEAYVKVDCSPEPQIMK-N 360
Qy 359 NEEAEDYDDDDTDSMDVRRDDDNPSGFIQIRSVAKKPKTWIHYIAAEBEDWDYAPLV 418
Db 361 NEEAEDYDDDDTDSMDVRRDDDNPSGFIQIRSVAKKPKTWIHYIAAEBEDWDYAPLV 418
Qy 419 LAPDRSFKSQYLNNGPORIGKVKVZFMAVDTSTETREAIQESGLIGPLLYGEVD 478
Db 419 PTDNGSKYSQYLSNGPRIGRKVKVZFMAVDTSTETREAIQESGLIGPLLYGEVD 478
Qy 479 TLLIIFKQASRPYNIYPHGIIDVAPLXSRALPKGVKELKOPPIIPGBIFKYKWTVTVED 538
Db 479 TLLIIFKQASRPYNIYPHGIIDVAPLXSRALPKGVKELKOPPIIPGBIFKYKWTVTVED 538
Qy 539 GPTKDPRLCTLYSSFNWMDRDLASGLIGPLLI CYKESVDOKGQVIXDKNVILPSVF 598
Db 539 GPTKDPRLCTLYSSFNWMDRDLASGLIGPLLI CYKESVDOKGQVIXDKNVILPSVF 598
Qy 599 DNRSWYLTENTORELPHAPAGVOLKEDPEQASINMHSINGYVDFSILQSLVCLHEVAYWII 658
Db 599 DNRSWYLTENTORELPHAPAGVOLKEDPEQASINMHSINGYVDFSILQSLVCLHEVAYWII 658
Qy 659 LSIGAOITDLSLFFSGYTPKHKVYEDTTLTLPFFSGEIVFMSNENPGLMILGCHNSDPRN 718
Db 659 LSIGAOITDLSLFFSGYTPKHKVYEDTTLTLPFFSGEIVFMSNENPGLMILGCHNSDPRN 718
Qy 719 RGMATLLKVSCKDKSTDIYEEIYEDIPQLNENNVNIDPERSFFQNTNHPNTRKKKPKDOS 778
Db 719 RGMATLLKVSCKDKSTDIYEEIYEDIPQLNENNVNIDPERSFFQNTNHPNTRKKKPKDOS 778
Qy 779 TTPENDIEKTDWFAHRTWPKLQNVSSDMLMLRQS-PTPGLSISLQEAKEYTFSD 837
Db 779 TTPENDIEKTDWFAHRTWPKLQNVSSDMLMLRQS-PTPGLSISLQEAKEYTFSD 837
Qy 838 DPSPGAIIDNSNLSMETHFRPQLEHSGDMVFTPBESGLQELNKLGTAAATELKKLDKVK 897
Db 838 DPSPGAIIDNSNLSMETHFRPQLEHSGDMVFTPBESGLQELNKLGTAAATELKKLDKVK 897
Qy 898 SSTSNNLI-STIPSDNLAAGTDNTSILGPPSPMPVHYDSQ-DTLTFLGKSSPLTESGGPLS 956
Db 898 SSTSNNLI-STIPSDNLAAGTDNTSILGPPSPMPVHYDSQ-DTLTFLGKSSPLTESGGPLS 956
Qy 957 LSEBNDKLLSGLMNSQESSGKNSVSTESGRFLFKGAAGPALLTKDNAJFKVSI 1016
Db 957 LSEBNDKLLSGLMNSQESSGKNSVSTESGRFLFKGAAGPALLTKDNAJFKVSI 1016
Qy 1017 LKNTKTSNNSATNRKTHDQSLIENSPSVWQN-ILSDETFKVTPLTHDRMLMDKNA 1075
Db 1017 LKNTKTSNNSATNRKTHDQSLIENSPSVWQN-ILSDETFKVTPLTHDRMLMDKNA 1075
Qy 1076 TALPILNHSNKTTSKKNWYQOKKEGPIPPDAQNDPMSPEKMLFLPESARWIORTHGKN 1135
Db 1076 TALPILNHSNKTTSKKNWYQOKKEGPIPPDAQNDPMSPEKMLFLPESARWIORTHGKN 1135
Qy 1136 SLNSGQSPKQVLSLGPXKVEGQNTFLSEKXKVVVVGKCEFTKDVGLKEMVFPSSRLFL 1195
Db 1136 SLNSGQSPKQVLSLGPXKVEGQNTFLSEKXKVVVVGKCEFTKDVGLKEMVFPSSRLFL 1195
Qy 1196 TNLNHLNHNTHNPKKTOEBIEKKEETLIIQENVVLPOHTHTGKNFMKNFLILSTRQNV 1255
Db 1196 TNLNHLNHNTHNPKKTOEBIEKKEETLIIQENVVLPOHTHTGKNFMKNFLILSTRQNV 1255

Db 1198 TTLSNVHENGHRHNOERNIOSEIE-KGALIEBKVVLPQVHEATGSKNELKDILILGTFRQNI 1256
Qy 1256 EGSYDGAAYAPVQDFRSLNDSNTRTKKHTAHFSK--KGEESNLEGLNQTKQIVKXACT 1313
Db 1257 --SLIYEVHVPVLQNIITSINNSTVQIHMEHFFKRRKDKETNBEGLVNKTRVWKVNY-- 1311
Qy 1314 TRISPNTSQNFVQSRKALKOPRLPLETELEKRIIVDDTSTOWSKNMKHLTPSLTQ 1373
Db 1312 -----PSQNIITQRKALQOFR-----STOWLKTINCSTOCTI 1349
Qy 1374 IDYNEKKGATQSPGLDCLTRSHSIPQANRSPPLIAKVSFPSPRIPIYIRVLFQDNSS 1433
Db 1350 IDHSEKWKKFIKSLSDS-SVAKSTQTNSSDSHIVKTSAPP---PIDLKRSPFQNKFS 1405
Qy 1434 HLPAAASY---RKXDSGVQESSHFLQAGKXNLSALITLMTDGDQREVSGLSGHSATNSV 1489
Db 1406 HVQASSYIYDFKTKSSRIQESNNFLKTKINPNSLAILPWNMPIDQGFKTSFGKSNINSV 1465
Qy 1490 TYKKVENVTLKPDLPKTSKGVILLIPKHVITQKDLFPETNSGSPGHLDLVEGSLLOQTE 1549
Db 1466 TYKKRENIIFLKPTLPESGKIELLPQVSIQEBILPFTSHGSPGHNLNKKVFLQXIQ 1525
Qy 1550 GAIKWNENRPGKVPFLRVATESAKTPSKLLDPLANDNHYGTQTPKDEWKSQBSPEKT 1609
Db 1526 GPTKWKAKRGE--SIKGT-ESSKNTSKLLNHHANDYHAAQIIPKDMWKSCKSPBII 1583
Qy 1610 AFKKKDTILS-NACESNHAIIAINEGONKPEIIVTNAQKQTERLCSQNPVPLXHQREI 1669
Db 1584 SIKQEDT-LSLRPFNGSHSIGA-NEKQNPQRETTWVQGTQRTCSQIPVPLKXHQREI 1642
Qy 1670 TRTILQSOQEBIDYDDTISIVEMKEDFDIYDEENQSPRSPOKTRHYFLAAVEHLWDYG 1729
Db 1643 --SAPQSEATDYDAITIE--TIEDPDIYSEDIKQSPRSPOKTRHYFLAAVEHLWDYG 1699
Qy 1730 MSSSHVILNRNAQSGSVQFKKVPFQETDGSFTQFLYRGELNEHLGLGPVIRAEVDN 1799
Db 1700 MSTS-HVLNRVQSDNVQFKKVFQETDGSFSQFLYRGELNEHLGLGPVIRAEVDN 1758
Qy 1790 IMVTFRNQASRPYSFYSSLSIYEBEORQGAERPKNPVKNETKYFWKVOHHMATAKDEF 1849
Db 1759 IMVTFRNQASRPYSFYSSLSIYEBEORQGAERPKNPVKNETKYFWKVOHHMATAKDEF 1817
Qy 1850 DCKAWAYFSDVDLEKDVHSLIGPLIIVCHTNTLNPAHSGOVTVQBFALPFTTFIDETKSWY 1909
Db 1818 DCKAWAYFSDVDLEKDVHSLIGPLIIVCHTNTLNPAHSGOVTVQBFALPFTTFIDETKSWY 1877
Qy 1910 FTENMERNCRAPONIOMEDPTFKENYRPHAINGYIMDTLPGLVMAQDQIRVYLLSMGN 1969
Db 1878 FTENMERNCRAPONIOMEDPTFKENYRPHAINGYIMDTLPGLVMAQDQIRVYLLSMGN 1937
Qy 1970 ENTHSHFGSHVPTVRKKEEYKXALNLYPGVFEIVEMLPKXAGIWEVCLIGELHAGM 2025
Db 1938 ENTHSHFGSHVPTVRKKEEYKXALNLYPGVFEIVEMLPKXAGIWEVCLIGELHAGM 1997
Qy 2030 STLFLVYKCOCTPLGWSAGHTRDFOITASGOYQWAPKIALRHYSGSINAWSTKPEFSW 2089
Db 1998 STLFLVYKCOCTPLGWSAGHTRDFOITASGOYQWAPKIALRHYSGSINAWSTKPEFSW 2057
Qy 2090 IKVDLLAPMIHGIKTQGAQKPESSYISQFIIMYSLDGKKWOTYRGNSTGTLMVFPGNV 2149
Db 2058 IKVDLLAPMIHGIKTQGAQKPESSYISQFIIMYSLDGKKWOTYRGNSTGTLMVFPGNV 2117
Qy 2150 DSGIKHNIPNPIIARIYIRLHETHYSIRSLAWELMGCDLNSCMPLGSKAISDAQI 2209
Db 2118 DSGIKHNIPNPIIARIYIRLHETHYSIRSLAWELMGCDLNSCMPLGSKAISDAQI 2177
Qy 2210 TASSYFTNMFATWSPKARLHLOCRSNAPVNNPKEWLQVDFQKTKMTVGTGTVQKVS 2269
Db 2178 TASSYFTNMFATWSPKARLHLOCRSNAPVNNPKEWLQVDFQKTKMTVGTGTVQKVS 2237
Qy 2270 LLTSMYKFEFLISSQDGHQWTLFFQNGKVKVQGNQDSTFPVNSLDDPILLIRYLRIHP 2329
Db 2238 LLTSMYKFEFLISSQDGHQWTLFFQNGKVKVQGNQDSTFPVNSLDDPILLIRYLRIHP 2297

Wed Apr 14 15:52:00 2004

np000123-328-355-581.rsp

```

QY 2330 QSWVHQIALRMEVLCGCHQADLY 2351
Db 2298 QIWEHQIALRLEILGCEAQOY 2319

RESULT 3
FAS_PIG STANDARD; PRT; 2133 AA.
AC P12263; Q95243;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Coagulation factor VIII precursor (Procoagulant component).
GN F8 OR CFB.
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX NCBI_TaxID=9823;
RN [1]
RP SEQUENCE FROM N.A.
RA Healey J.F., Lubin I.M., Lollar P.;
RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE OF 705-1573 FROM N.A.
RX MEDLINE=86287365; PubMed=3016730;
RA Toole J.J., Pittman D.D., Orr E.C., Murtha P., Wasley L.C.,
RA Kaufman R.J.;
RA "A large region (approximately equal to 95 kDa) of human factor VIII
RT is dispensable for in vitro procoagulant activity."
RL Proc. Natl. Acad. Sci. U.S.A. 83:5939-5942(1986).
RN [3]
RP SEQUENCE OF 392-759 FROM N.A.
RX MEDLINE=94179260; PubMed=7510695;
RA Lubin I.M., Healey J.F., Scandella D., Runge M.S., Lollar P.;
RT "Elimination of a major inhibitor epitope in factor VIII."
RL J. Biol. Chem. 269:8639-8641(1994).
CC -!- FUNCTION: Factor VIII, along with calcium and phospholipid, acts
CC as a cofactor for factor IXa when it converts factor X to the
CC activated form, factor Xa.
CC -!- SUBCELLULAR LOCATION: Extracellular.
CC -!- SIMILARITY: Contains 3 F5/8 type A domains.
CC -!- SIMILARITY: Contains 2 F5/8 type C domains.
CC -!- SIMILARITY: STRONG TO COAGULATION FACTOR V.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC -----
CC EMBL; U49517; AAB06705.1; .
CC DR PIR; A25945; A25945.
CC DR PIR; T42763; T42763.
CC DR HSP; P00451; 1CGF.
CC InterPro; IPR001117; Cu-oxidase.
CC InterPro; IPR008972; Cupredoxin.
CC InterPro; IPR000421; FAS8_C.
CC InterPro; IPR008979; Gal_bind_like.
CC Pfam; PF00754; F5_F8_typeC_2.
CC SMART; SM00231; FAS8C_2.
CC PROSITE; PS01285; FAS8C_1; 2.
CC PROSITE; PS01286; FAS8C_2; 2.
CC PROSITE; PS00022; FAS8C_3; 2.
CC PROSITE; PS00079; MULTICOPPER_OXIDASE1; 3.
KW blood coagulation; Repeat; plasma; Acute phase; Calcium;
KW Signal; Glycoprotein; Sulfation.
FT CHAIN 1 19 POTENTIAL.
FT CHAIN 20 2133 COAGULATION FACTOR VIII.
FT DOMAIN 20 357 F5/8 TYPE A 1.

```

```

20 199
207 357
399 730
399 730
583 730
760 1599
1495 1822
1495 1659
1569 1822
1822 1970
1975 2127
291 392
759 760
1449 1450
1490 1491
737 737
738 738
742 742
173 199
547 573
1633 1659
1822 1970
1975 2127
233 233
239 259
601 601
929 929
985 985
1025 1025
1111 1111
1181 1181
1208 1208
1245 1245
1265 1265
1335 1335
1408 1408
1611 1611
1919 1919
713 713
734 734
792 792
1133 1133
1191 1191
1209 1209
1437 1437
1456 1456
1539 1539
1546 1546
2133 2133
SQ SEQUENCE

Query Match 64.9%; Score 8057; DB 1; Length 2133;
Best Local Similarity 66.9%; Pred. No. 0;
Matches 1575; Conservative 209; Mismatches 345; Indels 226; Gaps 17;

QY 1 MQELSTCFPLCLLRFCFSATRRYYILGAVELSDYMQSD-LGELPVDARPPVPKSPFF 59
Db 1 MQELSTCFVFLCLPLGFSAIRYYILGAVELSDYRQSELRLHVDTRPPATAPGLPL 60
QY 60 NTSVYVYKTLFVEFTDHLFNIAKPRPPWMLGLPTTQAEVYDVVITLKNMASHPSLHA 119
Db 61 GPSVLYKTVFVEFTDQLFSVARPPRPPWMLGLPTTQAEVYDVVITLKNMASHPSLHA 120
QY 120 VGVSYKASGAEDDDQTSOREKEDDKVFGGSHYTVYQVLKENGWASDPLCLITYSYLS 179
Db 121 VGVSWKSSGEAEYEDHTSQREKEDDKVLPKGSQTVYVQVLKENGWASDPLCLITYSYLS 180
QY 180 HVDLVKOLNSGLIGALLVCREGSLAKETQTLKFKILLFAVFDGKSWHSETKNSLMQDR 239
Db 181 HVDLVKOLNSGLIGALLVCREGSLTRSTQNLEHVVLLFAVFDGKSWHSAENDSWTRAM 240
QY 240 DAASARAWPKCHTVNGYVNRSLPGLICCHRSVYVHWGMGTTPPEVHSIFLEGHTFLVRN 299
Db 241 DPAPARAPAMHTVNGYVNRSLPGLIGCHKSVYVHWGMGTSPVHASFIELEGHTFLVRH 300

```


Wed Apr 14 15:52:00 2004

np000123-328-355-581.rsp

DT 01-OCT-1989 (Rel. 12, Created)
 DT 01-FEB-1994 (Rel. 28, Last sequence update)
 DT 15-MAR-2004 (Rel. 43, Last annotation update)
 DE Coagulation factor V precursor (Activated protein C cofactor).
 GN F5.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A., AND VARIANT MET-1764.
 RX MEDLINE=92232668; PubMed=1567832;
 RA Cripe L.D., Moore K.D., Kane W.H.;
 RT "Structure of the gene for human coagulation factor V";
 RL Biochemistry 31:3777-3785(1992).
 RN [2]
 RP SEQUENCE FROM N.A., AND VARIANT GHJ-925.
 RX MEDLINE=87260886; PubMed=3110773;
 RA Jenny R.J., Pittman D.D., Toole J.J., Kriz R.W., Aldape R.A.,
 RA Hewick R.M., Kaufman R.J., Mann K.G.;
 RT "Complete cDNA and derived amino acid sequence of human factor V";
 RL Proc. Natl. Acad. Sci. U.S.A. 84:4846-4850(1987).
 RN [3]
 RP SEQUENCE OF 1-1600 FROM N.A., AND VARIANTS GIJ-925 AND ILE-1285.
 RX MEDLINE=88107560; PubMed=2827731;
 RA Kane W.H., Ichinose A., Hagen F.S., Davie E.W.;
 RT "Cloning of cDNAs coding for the heavy chain region and connecting
 RT region of human factor V, a blood coagulation factor with four types
 RT of internal repeats";
 RL Biochemistry 26:6508-6514(1987).
 RN [4]
 RP SEQUENCE OF 1188-1215 AND 1315-2224 FROM N.A.
 RX MEDLINE=86313665; PubMed=3032220;
 RA Kane W.H., Davie E.W.;
 RT "Cloning of a cDNA coding for human factor V, a blood coagulation
 RT factor homologous to factor VIII and ceruloplasmin";
 RL Proc. Natl. Acad. Sci. U.S.A. 83:6800-6804(1986).
 RN [5]
 RP PARTIAL SEQUENCE FROM N.A.
 RC TISSUE=Fibroblast;
 RX MEDLINE=93203619; PubMed=8454869;
 RA Shen N.L.L., Fan S.-T., Pyati J., Iapolla R.J.,
 RA Edgington T.S.;
 RT "The serine protease cofactor factor V is synthesized by
 RT lymphocytes";
 RL J. Immunol. 150:2992-3001(1993).
 RN [6]
 RP SULFATION.
 RX MEDLINE=94264012; PubMed=8204629;
 RA Pittman D.D., Tomkinson K.N., Michnick D., Selighsohn U.,
 RA Kaufman R.J.;
 RT "Posttranslational sulfation of factor V is required for efficient
 RT thrombin cleavage and activation and for full procoagulant activity.";
 RL Biochemistry 33:6952-6959(1994).
 RN [7]
 RP SULFATION
 RX MEDLINE=90366699; PubMed=2168225;
 RA Horton G.B.;
 RT "Sulfation of tyrosine residues in coagulation factor V";
 RL Blood 76:946-952(1990).
 RN [8]
 RP X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS) OF 2055-2224.
 RX MEDLINE=20052169; PubMed=1058886;
 RA Macedo-Ribeiro S., Bode W., Huber R., Quinn-Allen M.A., Kim S.W.,
 RA Ortel T.L., Bourenkov G.P., Bartunik H.D., Stubbs M.T., Kane W.H.,
 RA Fuentes-Prior P.;
 RT "Crystal structures of the membrane-binding C2 domain of human
 RT coagulation factor V";
 RL Nature 402:434-439(1999).
 RN [9]
 RP VARIANT MET-1764.
 RX MEDLINE=95179146; PubMed=7874144;
 RA Bayston T.A., Ireland H., Olds R.J., Thein S.L., Lane D.A.;

RT "A polymorphism in the human coagulation factor V gene.";
 RL Hum. Mol. Genet. 3:2085-2085(1994).
 RN [10]
 RP VARIANT APCR GLN-534.
 RX MEDLINE=94217810; PubMed=8164741;
 RA Bertina R.M., Koelmen B.P.C., Koster T., Rosendaal P.R.,
 RA Dirven R.J., de Ronde H., van der Velden P.A., Reitsma P.H.;
 RT "Mutation in blood coagulation factor V associated with resistance to
 RT activated protein C";
 RL Nature 369:64-67(1994).
 RN [11]
 RP VARIANTS ILE-1285 AND ARG-1327.
 RX MEDLINE=96351768; PubMed=8713778;
 RA Lunghi B., Iacoviello L., Gemmati D., Dilasio M.G., Castoldi E.,
 RA Pinotti M., Castaman G., Redaelli R., Mariani G., Marchetti G.,
 RA Bernardi F.;
 RT "Detection of new polymorphic markers in the factor V gene:
 RT association with factor V levels in plasma";
 RL Thromb. Haemost. 75:45-48(1996).
 RN [12]
 RP VARIANT APCR GLY-334, AND VARIANT LYS-513.
 RX MEDLINE=98122763; PubMed=9454741;
 RA Chan W.P., Lee C.K., Kwong Y.L., Lam C.K., Liang R.;
 RT "A novel mutation of Arg306 of factor V gene in Hong Kong Chinese";
 RL Blood 91:1135-1139(1998).
 RN [13]
 RP VARIANT APCR THR-334.
 RX MEDLINE=98122764; PubMed=9454742;
 RA Williamson D., Brown K., Luddington R., Baglin T.;
 RT "Factor V Cambridge: a new mutation (Arg306-to-Thr) associated with
 RT resistance to activated protein C";
 RL Blood 91:1140-1144(1998).
 RN [14]
 RP VARIANTS HIS-107; THR-413; LYS-513; SER-809; THR-817; ARG-858;
 RX ARG-865; GLU-925; GLN-1146; ALA-1530; SER-1685; VAL-1749; MET-1764;
 RP ILE-1820 AND GLY-2222, AND VARIANT APCR GLN-534.
 RX MEDLINE=99318093; PubMed=10391209;
 RA Cargill M., Altshuler D., Ireland J., Sklar P., Ardlie K., Patil N.,
 RA Shaw N., Lane C.R., Lim E.P., Kalyanaraman N., Nemesh J., Ziaugra L.,
 RA Friedland L., Rolfe A., Warrington J., Lipschutz R., Daley G.Q.,
 RA Lander E.S.;
 RT "Characterization of single-nucleotide polymorphisms in coding regions
 RT of human genes";
 RL Nat. Genet. 22:231-238(1999).
 RN [15]
 RP ERRATUM.
 RA Cargill M., Altshuler D., Ireland J., Sklar P., Ardlie K., Patil N.,
 RA Shaw N., Lane C.R., Lim E.P., Kalyanaraman N., Nemesh J., Ziaugra L.,
 RA Friedland L., Rolfe A., Warrington J., Lipschutz R., Daley G.Q.,
 RA Lander E.S.;
 RL Nat. Genet. 23:373-373(1999).
 RN [16]
 RP VARIANT APCR HIS-2102.
 RX MEDLINE=21847288; PubMed=1858490;
 RA Schrijver I., Houissa-Kastally R., Jones C.D., Garcia K.C.,
 RA Zehnder J.L.;
 RT "Novel factor V C2-domain mutation (R2074H) in two families with
 RT factor V deficiency and bleeding";
 RL Thromb. Haemost. 87:254-259(2002).
 CC -!- FUNCTION: Coagulation factor V is a cofactor that participates
 CC with factor Xa to activate prothrombin to thrombin.
 CC -!- SUBUNIT: Factor V is composed of a heavy chain and a light
 CC chain, noncovalently bound. The interaction between the two chains
 CC is calcium-dependent.
 CC -!- DOMAIN: Domain B contains 35 x 9 AA tandem repeats, and 2 x 17 AA
 CC repeats.
 CC -!- PTM: Thrombin activates factor V proteolytically to the active
 CC cofactor, factor Va (formation of a heavy chain at the N-
 CC terminus and a light chain at the C-terminus).
 CC -!- PTM: Sulfation is required for efficient thrombin cleavage and
 CC activation and for full procoagulant activity.
 CC -!- DISEASE: Defects in F5 are the cause of Owren parahemophilia
 CC [MIM:227400], an hemorrhagic diathesis.

CHAIN		1612	2258	(BY SIMILARITY).	CONJUGATION FACTOR V LIGHT CHAIN (BY SIMILARITY).
FT	DOMAIN	30	329	FS/8 TYPE A 1.	
FT	DOMAIN	30	193	PLASTOCYANIN-LIKE 1.	
FT	DOMAIN	203	329	PLASTOCYANIN-LIKE 2.	
FT	DOMAIN	348	683	FS/8 TYPE A 2.	
FT	DOMAIN	348	525	PLASTOCYANIN-LIKE 3.	
FT	DOMAIN	535	683	PLASTOCYANIN-LIKE 4.	
FT	DOMAIN	691	1611	B.	
FT	DOMAIN	1168	1539	41 X 9 AA APPROXIMATE TANDEM REPEATS OF T-L-S-P-D-L-[GS]-[HQ]-T.	
FT	REPEAT	1168	1176	1.	
FT	REPEAT	1177	1185	2.	
FT	REPEAT	1186	1194	3.	
FT	REPEAT	1195	1203	4.	
FT	REPEAT	1204	1212	5.	
FT	REPEAT	1213	1221	6.	
FT	REPEAT	1222	1230	7.	
FT	REPEAT	1231	1239	8.	
FT	REPEAT	1240	1248	9.	
FT	REPEAT	1249	1257	10.	
FT	REPEAT	1258	1266	11.	
FT	REPEAT	1267	1275	12.	
FT	REPEAT	1276	1284	13.	
FT	REPEAT	1285	1293	14.	
FT	REPEAT	1294	1302	15.	
FT	REPEAT	1303	1311	16.	
FT	REPEAT	1312	1320	17.	
FT	REPEAT	1321	1329	18.	
FT	REPEAT	1330	1338	19.	
FT	REPEAT	1339	1347	20.	
FT	REPEAT	1348	1356	21.	
FT	REPEAT	1357	1365	22.	
FT	REPEAT	1366	1374	23.	
FT	REPEAT	1375	1383	24.	
FT	REPEAT	1384	1392	25.	
FT	REPEAT	1393	1401	26.	
FT	REPEAT	1402	1410	27.	
FT	REPEAT	1411	1419	28.	
FT	REPEAT	1420	1428	29.	
FT	REPEAT	1429	1437	30.	
FT	REPEAT	1438	1446	31.	
FT	REPEAT	1447	1455	32.	
FT	REPEAT	1456	1464	33.	
FT	REPEAT	1465	1473	34.	
FT	REPEAT	1474	1482	35.	
FT	REPEAT	1483	1491	36.	
FT	REPEAT	1492	1500	37.	
FT	REPEAT	1501	1509	38.	
FT	REPEAT	1510	1518	39.	
FT	REPEAT	1519	1527	40.	
FT	REPEAT	1531	1539	FS/8 TYPE A 3.	
FT	DOMAIN	1616	1941	PLASTOCYANIN-LIKE 5.	
FT	DOMAIN	1616	1785	PLASTOCYANIN-LIKE 6.	
FT	DOMAIN	1795	1941	FS/8 TYPE C 1.	
FT	DOMAIN	1941	2095	FS/8 TYPE C 2.	
FT	DOMAIN	2100	2255	CLEAVAGE (BY THROMBIN) (BY SIMILARITY).	
FT	SITE	737	738	CLEAVAGE (BY THROMBIN) (BY SIMILARITY).	
FT	SITE	1029	1030	CLEAVAGE (BY THROMBIN) (BY SIMILARITY).	
FT	SITE	1611	1612	CLEAVAGE (BY THROMBIN) (BY SIMILARITY).	
FT	DISULFID	167	193	BY SIMILARITY.	
FT	DISULFID	248	329	BY SIMILARITY.	
FT	DISULFID	499	525	BY SIMILARITY.	
FT	DISULFID	602	683	BY SIMILARITY.	
FT	DISULFID	1759	1785	PROBABLE.	
FT	DISULFID	1941	2095	BY SIMILARITY.	
FT	DISULFID	2100	2255	BY SIMILARITY.	
FT	MOD_RES	692	696	SULFATION (POTENTIAL).	
FT	MOD_RES	696	696	SULFATION (POTENTIAL).	

FT	MOD_RES	724	724	SULFATION (POTENTIAL).	
FT	MOD_RES	726	726	SULFATION (POTENTIAL).	
FT	MOD_RES	745	745	SULFATION (POTENTIAL).	
FT	MOD_RES	1560	1560	SULFATION (POTENTIAL).	
FT	MOD_RES	1576	1576	SULFATION (POTENTIAL).	
FT	MOD_RES	1581	1581	SULFATION (POTENTIAL).	
FT	MOD_RES	1584	1584	SULFATION (POTENTIAL).	
FT	MOD_RES	1588	1588	SULFATION (POTENTIAL).	
FT	MOD_RES	1631	1631	SULFATION (POTENTIAL).	
FT	CARBOHYD	153	153	N-LINKED (GLCNAC. .)	
FT	CARBOHYD	225	225	N-LINKED (GLCNAC. .)	
FT	CARBOHYD	239	239	N-LINKED (GLCNAC. .)	
FT	CARBOHYD	297	297	N-LINKED (GLCNAC. .)	
FT	CARBOHYD	382	382	N-LINKED (GLCNAC. .)	
FT	CARBOHYD	460	460	N-LINKED (GLCNAC. .)	
FT	CARBOHYD	467	467	N-LINKED (GLCNAC. .)	
FT	CARBOHYD	553	553	N-LINKED (GLCNAC. .)	
FT	CARBOHYD	741	741	N-LINKED (GLCNAC. .)	
FT	CARBOHYD	752	752	N-LINKED (GLCNAC. .)	
FT	CARBOHYD	760	760	N-LINKED (GLCNAC. .)	
FT	CARBOHYD	776	776	N-LINKED (GLCNAC. .)	
FT	CARBOHYD	782	782	N-LINKED (GLCNAC. .)	
FT	CARBOHYD	899	899	N-LINKED (GLCNAC. .)	
FT	CARBOHYD	960	960	N-LINKED (GLCNAC. .)	
FT	CARBOHYD	1048	1048	N-LINKED (GLCNAC. .)	
FT	CARBOHYD	1057	1057	N-LINKED (GLCNAC. .)	
FT	CARBOHYD	1066	1066	N-LINKED (GLCNAC. .)	
FT	CARBOHYD	1073	1073	N-LINKED (GLCNAC. .)	
FT	CARBOHYD	1089	1089	N-LINKED (GLCNAC. .)	
FT	CARBOHYD	1174	1174	N-LINKED (GLCNAC. .)	
FT	CARBOHYD	1480	1480	N-LINKED (GLCNAC. .)	
FT	CARBOHYD	1537	1537	N-LINKED (GLCNAC. .)	
FT	CARBOHYD	1597	1597	N-LINKED (GLCNAC. .)	
FT	CARBOHYD	1737	1737	N-LINKED (GLCNAC. .)	
FT	CARBOHYD	1886	1886	N-LINKED (GLCNAC. .)	
FT	CARBOHYD	2044	2044	N-LINKED (GLCNAC. .)	
FT	CARBOHYD	2243	2243	N-LINKED (GLCNAC. .)	
SQ	SEQUENCE	2258	2258	N-LINKED (GLCNAC. .)	

Query Match 22.4%; Score 2776.5; DB 1; Length 2258;
Best Local Similarity 30.4%; Pred. No. 4.7e-135;
Matches 768; Conservative 388; Mismatches 874; Indels 497; Gaps 77;

QY	22	RYLYGAVELSWDMQSDLGELPVDARFPFRVPKFPFNTSVVYKTLFVEFDHLFNIA	81
Db	32	RQFYVAAQSIWNHDE-----PTHSSSPFRATS--FKIVYREY-EAYFQKE	76
QY	82	KGRPPWGLLGPITQAEVVDTVVITLKNMASHPVSLHAGVSYSKASEGARYDDQTSORE	141
Db	77	KPFSRMSGLLGPILYADVGDIMKVHFNKADKPLSIHPQGIKYKFAEGASYPDHTFLVE	136
QY	142	KEDDKVFPGSSHTYVMQVLKENGPMASDPLCTYSYLSHVDLVKDLNSGLIGALLVCREG	201
Db	137	KMDDAVAPQOEYIYEWNISDSGTHNDPPCLTHIYYSYENLIQDNFNSGLIGLLICKKG	196
QY	202	SLAKEKOTL--HKFILLFAVDEGKSWHSETKNSLNQDRDAASARAWPQOHVWNGVNR	259
Db	197	TLTEDIQKMFQYVIMFAVDFDESKSNQSSS-----LMYVNGVNG	240
QY	260	SLPLGLICRHSYVHVHVGITTPVEHSIFLEGHTFLVRNHRQASLEISPTLTATLL	319
Db	241	TMDITVCAYDHSIWHLLIGNSSGPEFSIHFSQVLEQNHKKVSATLTVSATSTANMTV	300
QY	320	MDLQGLLSCHISHHQDGMENAVYKVDSCPEEPOLINKNNEAEYDDDDITDSEMDVRF	379
Db	301	SPEGWPISSLIKHFOAGQCAVIDIKNCARTRKPKK-----LTRDQ-----	343
QY	380	DDNSPSFIORSVAKGPKTWVHYTAABEDWDYALVLPDDRYSKYSLNNGPORIG	439
Db	344	-----RRHIXWEYFIAABEVIWDYADIIIPANMDKKYRSHLDNFSNQIG	388
QY	440	RKYKVRFMVATDETFTKTR--EATQCHESGILGPLLYGEVGDTLIIIFKNQASFPYNYPH	497

389 KHKKVYKQKQDESFTKRLNPNKEDGILGPVIRAQVRDTLKVFNKMASRYSIYPH 448
 498 GIT-----DVRPYSRRLPKGVXHLKDFPLPGEIPKYKWTVVVBGPTKSDPRCLTRY 551
 449 GVTSPYEDDVNSSTSDNMTIR-----AVQGETVYKWNILSEDEPTENDACCLTRP 503
 552 YSSFNMRDLASGLIGLPLLYCKESVDOKGNOIMSDKRNVLSPVEDENRSWLTENIQ 611
 504 YYSNVDIETDIASGLIGLICKRSLDKRGLOQTADIEQKAVFAVDENKSWIEDNIY 563
 612 RFLPNZAGVOLDEPFOASNMHSGVVFDSL-QLSVCLHEVAYWYILS-GAOTDFLVS 670
 564 KPCNEPKVKDDPKFYESNMIMSTINGIVPESIPTLGFCFDDTVQWHFCSVRTHNULTI 623
 671 PFSQYFKHVMYEDTLPLFPSPGTFVMSMENPGLWILGCHNSDFRNGMTALLKVSSC 730
 624 HPTGHSFYKGRHEDTLPLFMGEHSVTVMNVTGMLTMTNSPRNKKQLKFRDVKC 683
 731 DNTG-DYVEDSEYDISAYLL-----SKNAIEP-----RSFQNS 765
 684 IRDDDESYEIIYEPSSSTTLTRKMDSEKKEBENDEYDQDLASVLGIRFSNS 743
 766 REPSTKOFNATTIP-ENDIEKTDFAFRCPMPKIQNVSSDMLMLRQSPTHGLSL 824
 744 LY--QEDDEFNLTALENNSEFIIP-----STD-----RAVDSNSSP 780
 825 SLOQEKYTFDD-----PSPGADSNNS-----LSEMTHERPOLHSGDMVFTP-BSG 873
 781 GNISRAPANTFEPKXILHPEATKAGSPRHTGLVKNLVLNRRRQHSOPYSEDIENP 840
 874 IQ-----LALNE-----KLGTAAATE--LKLDFKVSSTNNLISTIP 909
 841 LOSVITGILLSPTGTGRNREPKKPKFAGDQAKHFFSQMERPAHKTGRH----IS 896
 910 SDMLAAGTONTSLGPPSPMHHVDSOLDITLPGKSSPTSGGSLSEBN-----962
 897 QDN-----SSSSSMGPLE-----DLSDDLILERK-DPSTINGKWLHVSEKSEIVQDA 945
 963 DSKLESGLMNSQES-----SWGNKVSSTESGELFKGKRAHGPALLTKDNALFKVSIILLKT 1019
 946 DEDVAVNKLFPNNPONASRWGENIPTFNK-----HGKQRGHPIFVTRH-----KLQE 993
 1020 NKTSNNS-----ATNKTHIDGSLIENSPPVQWQIL-ESDTEFKKVTPLIHDRM 1069
 994 RODEGNILKKGRFFITRKKKERRKFWHVPSPSPNPLRGEANTFES-----1043
 1070 LMDXNATALRNLHNSKNTTSSKNMVMQKKEGIPDPAQN---PDMSFFKMLFLPSAR 1126
 1044 --DRQNHSLILHESNET-----FPPTDLNQTFPSMNLSLIASHPD---1082
 1127 WQRTHGKNSLN-SQOGSPKQVLSLGPESVQGNFLSEKNKVVVGKGFTHQVGLKM 1185
 1093 -----HNQNLFNTHQTSPP-----LDLYQT 1103
 1186 VPPSS--RNLPLTLNLDHNHNTENQEKKIOEIEKKEITLQENVLPQIHTVTKNFM 1243
 1104 VTPDEPYQAPLODLDPHSTAVPSHQSSLPEPI-----QME-----1140
 1244 KNLFILLSTRQNVGSDYDGAAPVLQDFRSLNDSNTRTKHTAHS---KKGEENLEG 1300
 1141 -----DYD-----LRNKASPTDVSENFSSFLKLAGHRTTSPDL- 1173
 1301 NOTKQIVEKYACT-----TBSINTSQNFVQTSKRALQFRLPLEETEL-----EK 1348
 1174 NOTLSPELSQTTLSDPGHVTLSPDLSQTTLSPLDLSHTLSP---DLGHTTSLDLSHT 1230
 1349 RIIVDDTSTQWKNMKHLPSTLTQIDVNEKEKGAITQSPDLSQTLRSHSIPOANRSP 1403
 1231 TSLDLSQTTLSPLDLSHTLS-----PDLGHTTSLDLSHTTSLDLSGHTLSP-D 1280
 1409 IAKVSFPSPIRPTLYTRVLQFQNSHLPAASRKKDSG-VQESSHFLQAKKNLJSLATL 1467
 1281 LSHHTLSPLDLSHTTSLDLSGHTTSLDLSGHTTSLDLSGHTTSLDLSGHTTSLDLSGHTTSL 1340

1468 T--LEMTGDQREYSGISGTSATNSVTYKKVENTVLP-----KPDLPKTS-----CKVEL 1513
 1341 SPDLSHHTLSPDLGHTTSLSPDL--SOT-----TSPDLGHTMTLSPDLSHHTLSPDLGHTTSL 1394
 1514 LPKVHIYQKDLPTFETNSGSPGHLDLVEGSILOQTEGAIKWNEANRPGKVPFLRVATPSS 1573
 1395 SP-----DLSHT-TLSPDLGHTMTLSPD--LGOTTLSPDLGHTTSLSPDLSHHTLSPDL 1444
 1574 AKTSPKLL-----DPLAW-----DNHYGTQIPKEEWSQSKSPKTAFFK 1613
 1445 HETLSPDLSPDLSPDLSPDLSPDLSPDLSPDLSPDLSPDLSPDLSPDLSPDLSPDLSPDL 1502
 1614 KDTLSLNACSNHAIANAINEGQNKPEIEVWAKQATERLCSQ-----NPVLAERHQ 1666
 1503 -----ISSSQSVTLPEFGQTSPPDLGQRPSPSPSHSTLNNTPTPREPNVWVUGLS 1554
 1667 RE-----TFTTLOSDQEB-----IDYDQTTISVEMKEDDFDIYDRENQSP-----1707
 1555 RDDGYVBIIPROQSESEEDYVKIDYVEDDPYQTDVTR-----DINSRPNDAAM 1608
 1708 --RSPOKXTRHYFAAVERLWDYGMSSPHVLRNRAQSGVPO---FKKVVFOETDGSF 1762
 1609 YLRNNGNRNVYJAAEELSMYDSK-----FTQREDIDVPEHTIYKXVVERKYLDSTF 1662
 1763 TOPLVRGELNHLGHLGPIYRAEVEDNIMVFRNQAASRPSFYSSLISYE-----ED 1814
 1663 TKLDPRGEVEEHLGILGPIIIRAEVDDVIQVRFNKLASRPSLHAGLSYKSSSEKTYED 1722
 1815 QFQGAEPKPNFKVKNETKYFWKVQHHVAPTDEFCKAWAYFSDVDFKDVHSLGIGL 1874
 1723 DSPEWFKEDNAVQPNASSIYVWHATERSGPSPGACAWAYAVSVPKDIHSLGIGL 1782
 1875 LVCHTNTLNPAGRGVTVQVEPALFTTIPDETYSKWFTENNERNCRAPCMIQMEDTFKEN 1934
 1783 LICCKTLHKENNPVDMREFFVLLFMVDFEKSSWYEEKFTRSWR-----LTSSEVKNS 1836
 1935 YRPHANGVIMDT--PGLVMAQOORIRWILLMSGNSNIHSIHFSGHFTVRKKEEYKMA 1994
 1837 HKPHANGMYN-LPGLRMYEOBWRHLHLNJJGSDIHHVPHGOTLLENGTQOHLQGV 1895
 1995 YNLYPGVEETVEMLPKAGIWRVECLIGELHAGMSTLFLVYSNKCOTPLGNASGHIRDP 2054
 1896 WPLLPGSPKTLNMTKSKAGWLLDTEVGENQORAGMOTPFLIIDRECKMPMGLSTGLIADS 1955
 2055 QITASQYQWAPKGLARLHYSGSINAWSTKEPS-----WIKVDLAEMLIHGKTQ 2107
 1956 QIKASEFWGHQPKLARLNNGSGYNAWIT-DKPSGESNSKPMIQQDMQREVVFTGTOTQ 2014
 2108 ARQKFSLYTSQFIIMVSLQKXWTVRGNSTGTLMVFFGNVDSGGIKHNIFFNPPIARY 2167
 2015 AKYLLKSYTYTFEFNVAISSQORNRWIFKGNSTKNVWYFNGNSDASTITENQFDPVVARY 2074
 2168 IRLPHYSIRSTILRMLMGCDLNSCNPLGMSKAISSAQITASSYFTNNFAT-WSPSK 2226
 2075 IRISPTSYNKPALRLLELQGEVNGCSTPLGMSGNIKNEQITASSFKSWGWDYWEFFR 2134
 2227 ARHLQGRSNARPOVNPXENLQVDFQKTMKVTVGTQGVKSLTSMYKKEFLISSQ 2286
 2135 ARLNAQGRVNAQWAKANNNNQWLQIDLLKTKITAITTQCGCKSLUSSEMYVRYTIQVSDR 2194
 2287 QHWTLPQNGKV--KYFOGNQDSFTPVVNSLDPPLLTRYLIRHPQSWVHOIALRMEYLG 2344
 2195 GVEKVSREKSSMVDKIFEGNNNKGHVKNFFNFPFIIISRFIRIIPKWNQSIARLRLBFG 2254
 2345 CEAOOLY 2351
 2255 C---DIY 2258

AC Q28107; Q28108;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Coagulation factor V precursor (Activated protein C cofactor).
GN F5.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RX MEDLINE=92147638; PubMed=1737753;
RA Guinto E.R., Esmen C.T., Mann K.G., Macgillivray R.T.;
RT "The complete cDNA sequence of bovine coagulation factor V.";
RL J. Biol. Chem. 267:2971-2978(1992).
RN [2]
RP DISULFIDE BONDS.
RX MEDLINE=95034740; PubMed=7947716;
RA xue J., Kalafatis M., Silveira J.R., Kung C., Mann K.G.;
RT "Determination of the disulfide bridges in factor Va heavy chain.";
RL Biochemistry 33:13109-13116(1994).
CC -!- FUNCTION: Coagulation factor V is a cofactor that participates
CC with factor Xa to activate prothrombin to thrombin.
CC -!- SUBUNIT: Factor Va is composed of a heavy chain and a light
CC chain, noncovalently bound. The interaction between the two chains
CC is calcium-dependent.
CC -!- DOMAIN: Domain B contains 29.5 X 9 AA tandem repeats, and 2 X 14
CC AA repeats.
CC -!- PTM: thrombin activates factor V proteolytically to the active
CC cofactor, factor Va (formation of a heavy chain at the N-
CC terminus and a light chain at the C-terminus).
CC -!- PTM: Sulfation is required for efficient thrombin cleavage and
CC activation and for full procoagulant activity (By similarity).
CC -!- SIMILARITY: Contains 3 F5/8 type A domains.
CC -!- SIMILARITY: Contains 2 F5/8 type C domains.
CC -!- SIMILARITY: STRONG, TO COAGULATION FACTOR VIII.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement. (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL, M81440; AAA30512.1; -;
CC EMBL, M81441; AAA30513.1; -;
CC FIR, A42580; KFE05.
CC HSP, P12259; ICDT.
CC InterPro: IPR001117; Cu-oxidase.
CC InterPro: IPR008972; Cupredoxin.
CC InterPro: IPR000421; FA58 C.
CC InterPro: IPR008979; Gal_bind_like.
CC Pfam: PF00394; Cu-oxidase_3.
CC Pfam: PF00754; F5_F8_type_C; 2.
CC SMART, SM00231; FA58C; 2.
CC PROSITE, PS01285; FA58C_1; 2.
CC PROSITE, PS01286; FA58C_2; 2.
CC PROSITE, PS50022; FA58C_3; 2.
CC PROSITE, PS00079; MULTICOPPER_OXIDASE1; 2.
CC Blood coagulation; Glycoprotein; Sulfation; Calcium; Signal; Zymogen;
KW Repeat.
FT SIGNAL 1 28 POTENTIAL.
FT CHAIN 29 2211 COAGULATION FACTOR V.
FT CHAIN 29 741 COAGULATION FACTOR HEAVY CHAIN (BY
FT SIMILARITY).
FT PROPEP 742 1564 ACTIVATION PEPTIDE (CONNECTING REGION)
FT (BY SIMILARITY).
FT CHAIN 1565 2211 COAGULATION FACTOR LIGHT CHAIN (BY
FT SIMILARITY).
FT

FT DOMAIN 30 327 F5/8 TYPE A 1.
FT DOMAIN 30 193 PLASTOCYANIN-LIKE 1.
FT DOMAIN 203 327 PLASTOCYANIN-LIKE 2.
FT DOMAIN 348 686 F5/8 TYPE A 2.
FT DOMAIN 348 525 PLASTOCYANIN-LIKE 3.
FT DOMAIN 535 686 PLASTOCYANIN-LIKE 4.
FT DOMAIN 696 1564 B.
FT DOMAIN 1124 1151 2 X 14 AA TANDEM REPEATS.
FT DOMAIN 1124 1137 1-1.
FT REPEAT 1124 1137 1-2.
FT REPEAT 1151 1151 1-1.
FT DOMAIN 1188 1453 30 X 9 AA APPROXIMATE TANDEM REPEATS OF
FT [AS]-L-S-P-D-[LP]-[GS]-Q-[TE].
FT REPEAT 1188 1196 2-1.
FT REPEAT 1197 1205 2-2.
FT REPEAT 1206 1214 2-3.
FT REPEAT 1215 1223 2-4.
FT REPEAT 1224 1232 2-5.
FT REPEAT 1233 1241 2-6.
FT REPEAT 1242 1250 2-7.
FT REPEAT 1251 1259 2-8.
FT REPEAT 1260 1268 2-9.
FT REPEAT 1269 1277 2-10.
FT REPEAT 1278 1286 2-11.
FT REPEAT 1287 1295 2-12.
FT REPEAT 1296 1304 2-13.
FT REPEAT 1305 1313 2-14.
FT REPEAT 1314 1322 2-15.
FT REPEAT 1323 1331 2-16.
FT REPEAT 1332 1340 2-17.
FT REPEAT 1341 1349 2-18.
FT REPEAT 1350 1358 2-19.
FT REPEAT 1359 1367 2-20.
FT REPEAT 1368 1376 2-21.
FT REPEAT 1377 1385 2-22.
FT REPEAT 1386 1394 2-23.
FT REPEAT 1395 1403 2-24.
FT REPEAT 1404 1412 2-25.
FT REPEAT 1413 1421 2-26.
FT REPEAT 1422 1430 2-27.
FT REPEAT 1431 1439 2-28.
FT REPEAT 1440 1444 2-29 (PARTIAL).
FT REPEAT 1445 1453 2-30.
FT DOMAIN 1569 1990 F5/8 TYPE A 3.
FT DOMAIN 1569 1738 PLASTOCYANIN-LIKE 5.
FT DOMAIN 1748 1890 PLASTOCYANIN-LIKE 6.
FT DOMAIN 1894 2048 F5/8 TYPE C 1.
FT DOMAIN 2053 2208 F5/8 TYPE C 2.
FT SITE 741 742 CLEAVAGE (BY THROMBIN) (BY SIMILARITY).
FT SITE 1034 1035 CLEAVAGE (BY THROMBIN) (BY SIMILARITY).
FT SITE 1564 1565 CLEAVAGE (BY THROMBIN).
FT DISULFID 167 193
FT DISULFID 248 329
FT DISULFID 499 525
FT DISULFID 607 688
FT DISULFID 1712 1738 PROBABLE.
FT DISULFID 1894 2048 BY SIMILARITY.
FT DISULFID 2053 2208 Sulfation (POTENTIAL).
FT MOD_RES 697 697 Sulfation (POTENTIAL).
FT MOD_RES 701 701 Sulfation (POTENTIAL).
FT MOD_RES 730 730 Sulfation (POTENTIAL).
FT MOD_RES 1513 1513 Sulfation (POTENTIAL).
FT MOD_RES 1529 1529 Sulfation (POTENTIAL).
FT MOD_RES 1537 1537 Sulfation (POTENTIAL).
FT MOD_RES 1541 1541 Sulfation (POTENTIAL).
FT CARBOHYD 225 225 N-LINKED (GLCNAC...)
FT CARBOHYD 239 239 N-LINKED (GLCNAC...)
FT CARBOHYD 297 297 N-LINKED (GLCNAC...)
FT CARBOHYD 382 382 N-LINKED (GLCNAC...)
FT CARBOHYD 460 460 N-LINKED (GLCNAC...)
FT CARBOHYD 553 553 N-LINKED (GLCNAC...)
FT CARBOHYD 587 587 N-LINKED (GLCNAC...)
FT CARBOHYD 745 745 N-LINKED (GLCNAC...)
FT CARBOHYD 756 756 N-LINKED (GLCNAC...)

FT	CARBOHYD	774	774	N-LINKED	(GLCNAC. .)	(POTENTIAL)	755
FT	CARBOHYD	780	780	N-LINKED	(GLCNAC. .)	(POTENTIAL)	755
FT	CARBOHYD	902	902	N-LINKED	(GLCNAC. .)	(POTENTIAL)	737
FT	CARBOHYD	952	952	N-LINKED	(GLCNAC. .)	(POTENTIAL)	84
FT	CARBOHYD	964	964	N-LINKED	(GLCNAC. .)	(POTENTIAL)	783
FT	CARBOHYD	1044	1044	N-LINKED	(GLCNAC. .)	(POTENTIAL)	783
FT	CARBOHYD	1053	1053	N-LINKED	(GLCNAC. .)	(POTENTIAL)	864
FT	CARBOHYD	1062	1062	N-LINKED	(GLCNAC. .)	(POTENTIAL)	838
FT	CARBOHYD	1071	1071	N-LINKED	(GLCNAC. .)	(POTENTIAL)	904
FT	CARBOHYD	1078	1078	N-LINKED	(GLCNAC. .)	(POTENTIAL)	898
FT	CARBOHYD	1094	1094	N-LINKED	(GLCNAC. .)	(POTENTIAL)	954
FT	CARBOHYD	1451	1451	N-LINKED	(GLCNAC. .)	(POTENTIAL)	942
FT	CARBOHYD	1490	1490	N-LINKED	(GLCNAC. .)	(POTENTIAL)	1004
FT	CARBOHYD	1500	1500	N-LINKED	(GLCNAC. .)	(POTENTIAL)	997
FT	CARBOHYD	1590	1590	N-LINKED	(GLCNAC. .)	(POTENTIAL)	1054
FT	CARBOHYD	1639	1639	N-LINKED	(GLCNAC. .)	(POTENTIAL)	1048
FT	CARBOHYD	1939	1939	N-LINKED	(GLCNAC. .)	(POTENTIAL)	1134
FT	CARBOHYD	1997	1997	N-LINKED	(GLCNAC. .)	(POTENTIAL)	1087
FT	CARBOHYD	2196	2196	N-LINKED	(GLCNAC. .)	(POTENTIAL)	1171
FT	VARIANT	587	592	NETLPA -> T (IN VARIANT 2)			1231
SQ	SEQUENCE	2211	AA; 248981 MW; CEBF9083738667C45 CRC64;				1139
Query Match 22.1%; Score 2743; DB 1; length 2211;							
Best Local Similarity 30.2%; Pred. No. 2.4e-133;							
Matches 764; Conservative 383; Mismatches 833; Indels 550; Gaps 80;							
QY	22	RYIYGLAVELSDWYQSDLCGLPVDARPPRVKSPDNTSVYKKTFLVBSFDHLENIA	81				
DB	32	ROFYAAQSIIRNYR-----PESTHL-----SSKPFETS--FKIVIREY-EAYFQKE	76				
QY	82	XRPWMLGGLGTIOAEVYDVITLKNWASHVPSLHVGVSYKASGAEYDDTOSORE	141				
DB	77	KQSRSTGLGLTILAEVGDIMKVKHAKHAPLSHAQGIKVSFEGASYSDETLPM	136				
QY	142	KEDDKVFPQGSITYVQVLEKNGPMASDPLCTYSYKSHVDLVKOLNSGLICALLVCREG	201				
DB	137	KMDAVAPQOEYTYEIISEHSGPTDDPPCLTHIYISVNLVFNLSGLIPELLICKRG	196				
QY	202	SLAKEKOTL--HKFTLLFAVDEGSHSEFKNSLUMQDRDAASARAWPKMHTVNGYVR	259				
DB	197	LTEDGTQKFKFQHVLMFAVDESNSMQTSS-----LMTVNGYVNG	240				
QY	260	SLPGLGCHRKSYVHVIGMTTPEVHSISPLEGHTFLVNRHQASLEISPTFLTAQTLL	319				
DB	241	TWPDITVCAHDHLSWHLIGMSGPELFSIHFNQVLEQNHKKISAITLVSAITSTANMV	300				
QY	320	MDLQGLLSCHSSHQDGMFAVVKVDSCEBEPOLIMKNEBAEDDDDLTJSEMVMVRF	379				
DB	301	SPEGRWTIASLIPRHFQAGMAYIDIKNCAKKTENPKK-----LTRDQ-----	343				
QY	380	DDNSFSFIQIRSVAKKHPTWVHYIAAEBEDWDYAPLAPDDRYSYKSOYLNNGPQRI	439				
DB	344	-----RRHKRWEYFIAAEVINDYAPIIPANNVKVRSILHLNFSNRIG	388				
QY	440	RKYKRVFYAYTETPKTR--EAIQHSGLIGLPLLYGEVGDITLLIPKQASAPYNIYH	497				
DB	389	KHYKVVYKQYQDSDFTKLEDPSSEGDGILPIIRAQVRDTLKIVFKNMASSYSIYH	448				
QY	498	GIT-----DVRPLYSRLPAGVKVXKDFPILPCEIFKYKWTVVEDGPTKSDPRCL	548				
DB	449	GVTFPSYDNBNVSSSTSGSNMTYRAVR-----PGETTYTKWNILESDEPIENDAOCL	500				
QY	549	TRYSSFPVNMERDLASGLIGPLTICYKESVDQKQNOIMSKRNVILFSVFDENRSWYLTE	608				
DB	501	TRPYSNVD--TRDLASGLIGLICKESLDRGIGQAADIEQQAFAVFDENKSWYIED	560				
QY	609	NICRPLNPAGVQLEDEFOASNTMH-----SINGYVDFSILQ--LSVCLHEVAVHYIISIG	662				
DB	561	NIYKFCENPEXVKRDDPKFYESNIMSNFTLPAINGVYVPESEIFILGFCPDDTVQVHFCSVG	620				
QY	663	ACTFELSVFSGYFKKMNVEDTLTLPFSGSEVFMENPGLWILGCHNSDFRNGMT	722				
DB	621	TQNDILATHTGHSFYKGRHEDTLTLPFMQGESVTVTMNVGFWMLTWMNSNPRSKLIR	680				

QY	723	ALLKSSCDKXTQDYVEDSYEDI-----SAYLLSKNA-----	755
DB	681	LRPRDAKTRNDDD--DSYEIIIFPSSGTAMTKKHDSSEIEDENADSDYQDELAII	737
QY	756	IEPRSFQNSRHSSTRQKQGNATIP--ENDIEKTDWFAHRTMPKIQNVSSSDLLMLLR	84
DB	738	LGERSFNSSL--NOEKDELNTALALEKDESEIPP--SANKRS--LDSNSS--	783
QY	815	QSPTPHGLSLSLDQEAKEYTFSDPSPGA-----IDSNNSLSMETHFRPOLHSG--	864
DB	784	RSVSRLLIAKFAESLKTLLHLEAPAGSPLHAGLDKNSALNP-----PWABHSSPYSE	838
QY	865	---DMVFTPESGLQJ---RLNEXLGTAAATELAKLDFKVSSTSNL	904
DB	839	DPREHPLSDVTGVSLFPFGTFKNRPAKHQRPQVGRQAKHFKSQTRPAKHITRL	898
QY	905	ISTIPSDNLAAAGTNTSS--LGPPSMPVHVDSDQDTTLFGKSS-----PLTESGGP	954
DB	899	-----SQNSSSRMGP-----WEDIPSDLLLLQCKDPFKILNGEHLVSEKGS	942
QY	955	LSLSEENNDSKLESGLMNSQES---SWGKNVSVSTSGRIKFKRAHGPALLT---	1004
DB	943	YEIIQDANENKTVAK--LENSPONDSTRWGENIPFKNS---HGKQSGHPTPLVTRKPLQ	997
QY	1005	---KDNALFKVS--SLUKTKNTSNNSATNRKTHIDGSLLENSPSVWQ-----NILES	1054
DB	998	DRQURRSRLKEGLFLIR-----TRKKKBEKPAHVHPLSPRSPHPLURGEVNASFS	1048
QY	1055	DTEPKKVTPLTHDMLMDKNATALRNHMSNKTSSKNMEMVQOKGPIFPDPAQNPDMS	1134
DB	1049	DRRN-----HSLLHASNSTLSID--LNQTFPSMNLSLA-----ASLPD--	1087
QY	1115	FFKMLFELPESARWQORTHGKNSLN--SQGSPSPQOLV--SLOPEKSVQCNFLSEKNKVV	1171
DB	1098	-----HDQTSNDTTSQTSPPDLYPTVSPSEHYQ-----	1117
QY	1172	KGKFTKGVKENVFPSSRNLFNLNLNHNHNTNQEKIQBIEKKTLLIQENVLP	1231
DB	1118	-----IFP-----IODSDPTHSTAPNSRSPD	1139
QY	1232	QIHTVTGKNFMKNLFLLLSTRONVEGSDYD---GAYAVLQDFRSL-----	1273
DB	1140	PHSTTAPNSRP-----PTQSQIPNVYDLNRAPITDVSQIFPSLELEVWQATSLDLS	1194
QY	1274	NDSNTRIKKHTAFAKKGPEENLEGNGOTKOIVKAYACTTRISNTSQONFVTOESKRA	1333
DB	1195	QFSISPDLCQMALSPDQESLSPLDQ-----TSLSPDLQESLSPLDQGTQ	1242
QY	1334	LQQFRPLPEET--ELEKRIYDDTSTQWSKNMKHLTP---STLTQIDYNEKEKGAITQ	1386
DB	1243	LSPD--PSQESLSPLDQGTALSPDPS-----QESLSPLDQGTALSPDPQESLSPLDQ	1294
QY	1387	SPLSDCITR---SHSIFQANRSPPLIAKVSSPPSPRIY--TRVLFDNSSHU--PAASRYKK	1443
DB	1295	TSLSPDLQESLSPLDQGTALSPDP-----SQESLSPLDQGTALSPD-----PSQESLS	1344
QY	1444	DSGVQSSSHFLQAKKNLSLAILTLEMTGDQREVSLGTSATNSVYKKTVENTVLPKPD	1503
DB	1345	DLQGTSLSPDL--QGESLSPLDQGTALSPDPSESL-----SPD	1381
QY	1504	LPKTS-----GKVELLPKVHIIYQKDLFPPTETNSGSPGHLDLVEGSLLOQTGAIKWNEAN	1558
DB	1382	LGQTSLSPLDQESLSP-----DLQGTALSP-----DLQESLSPLD-----	1418
QY	1559	RPKVPPLRVATESAKTFSLKLDPLAWNHYGTQIPKBEWKSQESKPEKTAFFKXDTI--	1617
DB	1419	--QGTPL-----SPDLSLESL-----SPDLSQDLKQTS	1446
QY	1618	--LSIN-----ACENHAIANEQONPETEVTWAKQGRTERLCSQ-----NPPVLKRR	1665
DB	1447	PLDINQTSHTSSQSLPLPEFCQTFPNADIGQPPPPDSTLNTFIPEEFLPLVVUGL	1506
QY	1666	QRE-----ITRTILQSDQEE-----IDYDOTISVEMKEDFDIYDENQSP-----	1707

Db 1507 SRDGGYIEIIPKQGESSEDEGEFFVAYNDPQYDRLT-----DINSRNPNDIAA 1560
 QY 1708 ---RSQKTRHYFIAAVERLWDYGMSSSHVLNRASQSS---VQO---FKVWVQEF 1758
 Db 1561 WYLSRNTGNKYYIAAEISWDYS-----KFVQSDVDVYVPEDTYKVKVFKYL 1611
 QY 1759 DG8PTQLYRGELNEHLLGLPIYRAVEDNIMVTPNQASRPVSFYSSLSIYE----- 1812
 Db 1612 DSTFTKLDPOCEVEZEHGILGPVIRAEVDDVIQVRFNLASRPVSLHAGLSVEKSSGK 1671
 QY 1913 --EPQCGAEPKPFVKNFTYFKVQVHMATPKDFCKAWAYFSDVDLXDVHSG 1870
 Db 1672 TYEDDSSEWFKEDNAIQNKTYTWHATRSFGENPGSACRAMAYISAVNPKDIHSG 1731
 QY 1871 IGPLLVCHTNTLPAHQGRQVTVQBFALFTFTIDFTKSWYFTENNERNCRAPCINQMDPT 1930
 Db 1732 IGPLLICRKGTLDETNMPVDMREDFVLLFVDFEDBKSWYDKPTSRWRASS-----E 1785
 QY 1931 FKENYRFAINGYNDITPLGLVMAQDQIRWYLLSMGSENIHSHTRSGHVTVRKKEEY 1990
 Db 1786 VKNSHEFFAINGMIYN-LPGLRMYDQEWVRHLNMLGGSRDLHVHFGQVLLNGTOQH 1844
 QY 1991 KVALNLYPGVPEVTEMLPSKAGIWRVCELGHEHLHAGMSTLFLVYKNCQTPPLGMASGH 2050
 Db 1945 QLGWVPLLPQGFKTELEKASPGWLLDTEVGEIQACGQTPFLIVDRECKMPGLSTGL 1904
 QY 2051 IRDFOITASQYQGWAPKLARLHYSGSINAW-----STK-EPFSWIKVDLLAPMIHGIK 2104
 Db 1905 IADSQIQASEFQWYBEPKRLARLNGSGSYNAWTAELSTEFNPEPMIQVDMQKEVLLTG 1964
 QY 2105 TQGAQKTSLSIYSFIIMYSLDGKKQWYRGNSGTGLMVFPNGVDSGKIKINFPPII 2164
 Db 1965 TQGAHLYKPYTTEFCVAISLDKRNRIKFNKSTRNVMVFGNSDASTIKENQIDPPV 2024
 QY 2165 ASYIRLHPHYSTIRSLRMLMGCDLNSGMLPGMESKAISSDAQITPASSYFTNMFAT-WS 2223
 Db 2025 ARYIIRISPTGSYNKPAIRLELQCEVNGCSTPLGMESGKIENKQITASSFPKKSWMGNYWE 2084
 QY 2224 PSKARLHLOGRNASPQVNNKPEWLOVDPQKTMKVTGTTQGVKSLTSMYVKEPLISS 2283
 Db 2085 PFLARLNAQGRVNAWQAKANNQWLDLKKITAIVTQCKSLSESMYVYKTYIH 2144
 QY 2284 SDQCHQWTLFPQNGKV--KVFQGNQDSFPPVNSLDPPLLTTRYLRHPQSWHQAIRME 2341
 Db 2145 SDQGTDKPKYREKSMWVKLFEKGNVNRGVKVFPPPIISRIRIIPKTNQOSIALRUE 2204
 QY 2342 VLGCBAQDLY 2351
 Db 2205 LFQC---DMY 2211
 RESULT 7
 CERU HUMAN STANDARD; PRT; 1065 AA.
 AC P0450; Q14063;
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 13-AUG-1987 (Rel. 05, Last sequence update)
 DT 15-MAR-2004 (Rel. 43, Last annotation update)
 DE Ceruloplasmin precursor (EC 1.16.3.1) (Ferroxidase).
 GN CP.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=86259737; PubMed=2873574;
 RA Koschinsky M.L., Funk W.D., van Oost B.A., McGallivray R.T.A.;
 RT "Complete cDNA sequence of human ceruloplasmin";
 RL Proc. Natl. Acad. Sci. U.S.A. 83:5086-5090(1986).
 RN [2]
 RP SEQUENCE OF 1-1006 FROM N.A.

RX MEDLINE=95017183; PubMed=7702601;
 RA Daimon M., Yamatani K., Igarashi M., Fukase N., Kawanami T.,
 RA Kato T., Tomimaga M., Sasaki H.;
 RT "Fine structure of human ceruloplasmin gene";
 RL Biochem. Biophys. Res. Commun. 208:1028-1035(1995).
 RN [3]
 RP SEQUENCE OF 1-40; 549-599; 784-829 AND 919-952 FROM N.A.
 RX MEDLINE=86275241; PubMed=3755405;
 RA Mercer J.F.B., Grimes A.;
 RT "Isolation of a human ceruloplasmin cDNA clone that includes the N-
 RT terminal leader sequence";
 RL FEBS Lett. 203:185-190(1986).
 RN [4]
 RP SEQUENCE OF 218-1065 FROM N.A.
 RX MEDLINE=86205876; PubMed=3486416;
 RA Yang F., Navlior S.L., Lum J.B., Cutshaw S., McCombs J.L.,
 RA Neberhaus K.H., McGill J.R., Adrian G.S., Moore C.M., Barnett D.R.,
 RA Bowman B.H.;
 RT "Characterization, mapping, and expression of the human ceruloplasmin
 RT gene";
 RL Proc. Natl. Acad. Sci. U.S.A. 83:3257-3261(1986).
 RN [5]
 RP SEQUENCE OF 20-1065.
 RX MEDLINE=84113493; PubMed=6582496;
 RA Takahashi N., Ortel T.L., Putnam F.W.;
 RT "Single-chain structure of human ceruloplasmin: the complete amino
 RT acid sequence of the whole molecule";
 RL Proc. Natl. Acad. Sci. U.S.A. 81:390-394(1984).
 RN [6]
 RP SEQUENCE OF 158-333; 519-724 AND 858-1065.
 RX MEDLINE=89117800; PubMed=6571985;
 RA Takahashi N., Bauman R.A., Ortel T.L., Dwulet F.E., Wang C.-C.,
 RA Putnam F.W.;
 RT "Internal triplication in the structure of human ceruloplasmin";
 RL Proc. Natl. Acad. Sci. U.S.A. 80:115-119(1983).
 RN [7]
 RP SEQUENCE OF 501-905.
 RX MEDLINE=81199407; PubMed=6940148;
 RA Dwulet F.E., Putnam F.W.;
 RT "Complete amino acid sequence of a 50,000-dalton fragment of human
 RT ceruloplasmin";
 RL Proc. Natl. Acad. Sci. U.S.A. 78:790-794(1981).
 RN [8]
 RP SEQUENCE OF 907-1065.
 RX MEDLINE=80137543; PubMed=6987229;
 RA Kingston I.B., Kingston B.L., Putnam F.W.;
 RT "Primary structure of a histidine-rich proteolytic fragment of human
 RT ceruloplasmin. I. Amino acid sequence of the cyanogen bromide
 RT peptides";
 RL J. Biol. Chem. 255:2886-2896(1980).
 RN [9]
 RP SEQUENCE OF 907-1065.
 RX MEDLINE=80137544; PubMed=6987230;
 RA Kingston I.B., Kingston B.L., Putnam F.W.;
 RT "Primary structure of a histidine-rich proteolytic fragment of human
 RT ceruloplasmin. II. Amino acid sequence of the tryptic peptides";
 RL J. Biol. Chem. 255:2886-2896(1980).
 RN [10]
 RP SEQUENCE OF 1007-1061 FROM N.A.
 RX MEDLINE=90285218; PubMed=2355023;
 RA Yang F.M., Friedrichs W.E., Cupples R.L., Barifacio M.J.,
 RA Sanford J.A., Horton W.A., Bowman B.H.;
 RT "Human ceruloplasmin. Tissue-specific expression of transcripts
 RT produced by alternative splicing";
 RL J. Biol. Chem. 265:10780-10785(1990).
 RN [11]
 RP REVIEW
 RX MEDLINE=22049919; PubMed=12055353;
 RA Hellman N.E., Gitlin J.D.;
 RT "Ceruloplasmin metabolism and function";
 RL Annu. Rev. Nutr. 22:439-458(2002).
 RN [12]
 RP X-RAY CRYSTALLOGRAPHY (3.1 ANGSTROMS).

RA Zaitseva I., Zaitsev V., Card G., Moshkov K., Bax B., Ralph A.,
RA Lindley P.,
RT "The X-ray structure of human serum ceruloplasmin at 3.1 A: nature of
the copper centres";
RL J. Biol. Inorg. Chem. 1:15-23 (1996).
CC -!- FUNCTION: Ceruloplasmin is a blue, copper-binding (6-7 atoms per
molecule) glycoprotein found in plasma. Four possible functions
are ferroxidase activity, amine oxidase activity, copper transport
and homeostasis, and superoxide dismutase activity.
CC -!- CATALYTIC ACTIVITY: 4 Fe(2+) + 4 H(+) + O(2) = 4 Fe(3+) + 2 H(2)O.
CC -!- COFACTOR: Binds 6 Cu-ions per molecule. This protein belongs to
the multicopper oxidases which contain three distinct Cu centers
known as type 1 or blue, type 2 or normal, and type 3 or coupled
binuclear.
CC -!- TISSUE SPECIFICITY: Expressed by the liver and secreted in plasma.
CC -!- DISEASE: Defects in CP are the cause of aceruloplasminemia
(MIM:604290). It is an autosomal recessive disorder of iron
metabolism characterized by iron accumulation in the brain as well
as visceral organs. Clinical features consist of the triad of
retinal degeneration, diabetes mellitus and neurological
disturbances.
CC -!- DISEASE: Ceruloplasmin levels are decreased in Wilson's disease,
in which copper cannot be incorporated into ceruloplasmin in
liver because of defects in the copper-transporting ATPase 2.
CC -!- SIMILARITY: Contains 3 F5/8 type A domains.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation
at the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M13699; AAA51976.1; -;
DR EMBL; D45045; BAA08085.1; -;
DR EMBL; D45044; BAA08084.1; -;
DR EMBL; D45028; BAA08084.1; JOINED.
DR EMBL; D45029; BAA08084.1; JOINED.
DR EMBL; D45030; BAA08084.1; JOINED.
DR EMBL; D45031; BAA08084.1; JOINED.
DR EMBL; D45032; BAA08084.1; JOINED.
DR EMBL; D45033; BAA08084.1; JOINED.
DR EMBL; D45034; BAA08084.1; JOINED.
DR EMBL; D45035; BAA08084.1; JOINED.
DR EMBL; D45036; BAA08084.1; JOINED.
DR EMBL; D45037; BAA08084.1; JOINED.
DR EMBL; D45038; BAA08084.1; JOINED.
DR EMBL; D45039; BAA08084.1; JOINED.
DR EMBL; D45040; BAA08084.1; JOINED.
DR EMBL; D45041; BAA08084.1; JOINED.
DR EMBL; D45042; BAA08084.1; JOINED.
DR EMBL; D45043; BAA08084.1; JOINED.
DR EMBL; D00025; BAA00019.1; -;
DR EMBL; X04135; CAA27752.1; -;
DR EMBL; X04136; CAA27753.1; -;
DR EMBL; X04137; CAA27754.1; -;
DR EMBL; X04138; CAA27755.1; -;
DR EMBL; M13536; AAA51975.1; -;
DR EMBL; J05506; -; NOT ANNOTATED_CDS.
DR PIR; A25443; KUFU.
DR PDB; 1KCM; 12-FEB-97.
DR GlycoSuiteDB; P00450; -;
DR SWISS-2DPAGE; P00450; HUMAN.
DR Sienra-2DPAGE; P00450; -;
DR GeneW; HGNC:2295; CP.
DR MIM; 117700; -;
DR MIM; 604290; -;
DR GO; GO:0005615; Extracellular space; TAS.
DR GO; GO:0004322; Ferroxidase activity; TAS.
DR GO; GO:0006879; Pi:iron ion homeostasis; TAS.
DR InterPro; IPR001117; Cu-oxidase.
DR InterPro; IPR002355; Cu_ox_copper_BS.

DR InterPro; IPR008972; Cupredoxin.
DR Pfam; PF00394; Cu-oxidase; 3.
DR PROSITE; PS00079; MULTICOPPER OXIDASE1; 3.
DR PROSITE; PS00080; MULTICOPPER OXIDASE2; 1.
KW Transport; Ion transport; Copper transport; Oxidoreductase; Plasma;
KW Metal-binding; Copper; Repeat; Signal; Glycoprotein; Polymorphism;
KW 3D-structure.
FT SIGNAL 1 19
FT CHAIN 20 1065 CERULOPLASMIN.
FT DOMAIN 20 357 F5/8 TYPE A 1.
FT DOMAIN 20 200 PLASTOCYANIN-LIKE 1.
FT DOMAIN 209 357 PLASTOCYANIN-LIKE 2.
FT DOMAIN 370 718 F5/8 TYPE A 2.
FT DOMAIN 370 560 PLASTOCYANIN-LIKE 3.
FT DOMAIN 370 718 PLASTOCYANIN-LIKE 4.
FT DOMAIN 570 1061 F5/8 TYPE A 3.
FT DOMAIN 730 900 PLASTOCYANIN-LIKE 5.
FT DOMAIN 908 1061 PLASTOCYANIN-LIKE 6.
FT CARBOHYD 138 136 N-LINKED (GLCNAC. . .).
FT CARBOHYD 358 358 N-LINKED (GLCNAC. . .).
FT CARBOHYD 397 397 N-LINKED (GLCNAC. . .).
FT CARBOHYD 762 762 N-LINKED (GLCNAC. . .).
FT DISULFID 174 200 PROBABLE.
FT DISULFID 276 357 PROBABLE.
FT DISULFID 534 560 PROBABLE.
FT DISULFID 637 718 PROBABLE.
FT METAL 120 120 COPPER (TYPE 2) (BY SIMILARITY).
FT METAL 122 122 COPPER (TYPE 3) (BY SIMILARITY).
FT METAL 180 180 COPPER (TYPE 3) (BY SIMILARITY).
FT METAL 182 182 COPPER (TYPE 1) (BY SIMILARITY).
FT METAL 994 994 COPPER (TYPE 2) (BY SIMILARITY).
FT METAL 997 997 COPPER (TYPE 3) (BY SIMILARITY).
FT METAL 999 999 COPPER (TYPE 3) (BY SIMILARITY).
FT METAL 1039 1039 COPPER (TYPE 3) (BY SIMILARITY).

Query Match 10.6%; Score 1322; DB 1; Length 1065;
Best Local Similarity 19.2%; Pred. No. 1.6e-60;
Matches 398; Conservative 203; Mismatches 417; Indels 1054; Gaps 29;

QY 5 LSTCFPLCLLRFCFSATRYLYGAVELSDVMQSLDGE---LPVAFPPRPVKPFPPNT 61
DB 6 LGIFLFLCSTP-AWAKEKHVYIGIETTWYD-ASDHGEKKLISVDEHSHNIYLONGPDR 63
QY 62 SVVYKTLFVEPTDHLFNIAKPRPPWMLGLGTIOAEVYDTVVITKMAHSPVSHVANG 121
DB 64 GRLYKKALYLQYTDFTFRITIEKPVMLGFLGIIKAETGDKVYVHLKMLASRPYTHSHG 123
QY 122 VSYWKASEGAEYDQOTSQREKEDDKVPFGSGSHYVWQVLKENGPMASDPLCLTYSLSHV 181
DB 124 ITYKSHGAIYPDNTIDFQRAADKVPGEQYTYMLLATEEQSPGEGDGNCTRIYHSHI 183
QY 182 DLVKDLNSGLIGALLVCREGLAKERTQTU-HKPIILLFAVDFDEGKSHSE-----TKNS 234
DB 184 DAKPDIASGLIGLIIICKDLSLCKEKHIDREFVVMFVVDENFNFWDENIKTYCSEP 243
QY 235 LMDRDRAASARAPKMHVTNGYVNSRSLPGLIGCHRSKSVYHVIQMGTTPEVHSIFLEGT 294
DB 244 EKVDKDNEDFQENRMYSVNGYTFGSLPLGLSMCAEDRKVYLPFGMGNEVDVHAAFFHQA 303
QY 295 FLVNRHQASLEISPIITFLTAQTLMDLQGLLSCHISSHQHDGMEAYVVKVDSCEPPEQL 354
DB 304 LTNKNYRIDTINLFPALTFDAYVVAQNPGEWMLSCQNLNHLKAGLQAFPOVQEC----- 357
QY 355 IMKNNEEAEDYDDDLTDSMDVVRFDNDSFPTQIRSVAKKPKTWWHIAAEEDWDY 414
DB 358 ---NKSSSKD-----NIRGKEVRH-----YVIAAEIWNV 385
QY 415 APL-----VLAPDDRYSKSOYLANGPQIRGRKYKVRFMAYTDTF---KTREAIQ 462
DB 386 APSGIDIFTKENTLAPGSDS--AVFFQGTTRIGSYKGVRYREYTTDASFNRKERGPEE 443
QY 463 HESGILGPLLYGEGVDTLLIIFKQASRPVNYIPEGI-----TDVRPLY---SRRLP 511

Db 444 EHLGILGEVIAEVDGDIRVTFHFKGAYPLSIPIGVRFKNGEITYSYNYPQSRSP 503
QY 512 KGVKHLKDFILPGEIEFKYKWTIVVEDGPTKSDPRC:TRYYSFVNMERDLASGLIGPLL 571
Db 504 PSASH-----VAPTETTYEWTVPKEVGPINADPVCLAKXYSAVDPTKDIFTGLIGPMK 558
QY 572 ICYKESVDQKQNSDKRNVLFSVDENRNSWLTENIQBFLENPAGVOLEDEPEQASN 631
Db 559 ICKGSLHANGKQKDVDFEFLPTVFDENESLLEDNIENETAPDQVQKXEDDFQESN 618
QY 632 IMHSGVWVDSLO-LSVCLHEVAVWYLSIGATDF:SVFPSCYTFKHKMVMYEDTLTLF 690
Db 619 KMSNGMGYQNGQLNCKGDSVVWVLFSGNADVHGIFYSGNTYLRGERRDTANLF 678
QY 691 PFSGETVMSMENFGLWILGCHNSDFNRGWTALLKVSSCDKNTGDDYEDYSAYLL 750
Db 679 POTSILHMFDETEGTFNVECLTTDHYTGMKQKYTVNQ----- 718
QY 751 SKNAIEPRFSQNSRHPSTKQFNATTIPENDIEKTDPFPAHRTPMKIQNVSSDILL 810
Db 719 ----- 718
QY 811 MLLQSPTPHGLSLDLOEAKYETFSDDPSGAIDSNNSLGEHTFRPQLHHSQDVFTP 870
Db 719 --ROSE-----DSTYLGERTY----- 735
QY 871 ESGLOLR-NEKIGTTAAETLKKLDFKVSSTNNLSTIPSNLAAGTNTSSLGPPSPV 930
Db 736 -----IAAVEVE----- 742
QY 931 HYDSOLDTTLFGKSSPITSEGGPLSLENNDSKILSGLMNSOESSGKNVSTESGR 990
Db 743 ----- 742
QY 991 LFKGKRAHPALLTKONALFKVISILKTKNTSNNATNRKTHIDGPSILLIENSPVMQN 1050
Db 743 -----WD- 744
QY 1051 ILESDETFKVTPLIHDRMLMDKNATALRLNMSNKTSSKNMVMQKKEGPIFEDAON 1110
Db 745 ----- 744
QY 1111 PMSFFKFLPESARWIOPTHKNSLSGQSPKQLVSLGPEKSVQGNFLSEKNVV 1170
Db 745 ----- 744
QY 1171 VGKGEFTKDVGLKEMVFPSSRNLFITNLDLHNENHNOEKKIOEBIEKKTILIQENVVL 1230
Db 745 ----- 744
QY 1231 POIHTVTGTFKFNFLJLSTRONVEGSDGAYAPVLQDFRSLNDSNTNRKHTAHFSKK 1290
Db 745 -----YSP----- 747
QY 1291 GBEENLEGLGNQTKQVEKYACTRISPNTSQGNFVTOQRKALKQFRLPLETELEKRI 1350
Db 748 -----QR----- 749
QY 1351 IVDDTSTQWSKNMKHLTPSLTIQIDYNEKEGAIQSPISDCLTRSHSIPQANRSPLEPIA 1410
Db 750 -----EWEKELHLJ----- 758
QY 1411 KVSSFPISIRPIYLRVLFDQNSSHLPAASVYRKXDSGVQESSHFLQGNKKNLSIALITLE 1470
Db 759 -----QEONVSNATL----- 768
QY 1471 MTGQREVGSIGTSATNSVYKVENIVLPKPDLPKTSKGVELLPKVHIYKOLFPPTETS 1530
Db 769 ----- 768
QY 1531 NGSPGHLDLVEGSLLOCTEGAIKWNEARFKVFFLRVATESSAKTFSKLDPLANDNHY 1590

Db 769 ----- 768
QY 1591 GTQIPXEENKSOEKSPEKTAFAKKKDTILSLNACSNHAIANAINEGQNKPSIEVTWAKQR 1650
Db 769 ----- 768
QY 1651 TERLCSQNPVLKRHOREITRITLQSQOEIDYDDTISVEMKKEDFIYDEDENQSPRSF 1710
Db 769 -----DKGE----- 772
QY 1711 QKTRHYFIAAVERLWDYGMSSPHVLNRAQSGSVPOFKKVVFOBRTDGSFTQPLRYGE 1770
Db 773 -----FYIGS-----KYKKVTVRQTDSTFRVPVERKA 800
QY 1771 LNEHLGILLGIVIRAEVEDNIMVTPRNOASRPYGFYSSLSIYSEEDQRCQCAPRKNFVK--- 1827
Db 801 EEHLGILGQLHADVGDKVLIKFNWATRPYSIHA-----HGQVETESTVTPTL 850
QY 1828 PNETKTYFWKVQHMAFTKDBFCKAWAYFSDVLEKDVHSGLIGLPLLVCHTNTLNPAHG 1887
Db 851 PGEITVYVWKIPERSGAGTEDSACIPWAYYSTVDQVLDLYSGLIGLPLVCRPYLKVFPN 910
QY 1888 RQVTVQBPALFFTFIDETKSNYFTENMERNCRAPCNIQMEDPTKENYRPHAIYIMDT 1947
Db 911 RRKL--EPALFLYFDENESWYLDNKTYSDDHEKVKDKDBEFIESKMAHNGRMPGN 968
QY 1948 LPGLVMAQDQIRWYLLSMGSNENIHSHFSGHVTVTKKEEYKVALYNLYPGVPETVEM 2007
Db 969 LQGTMHVGDENVWYLMGMGNEIDLHTVHFHGHFQYKRGVYSSDVFDPFGTYQTLEM 1028
QY 2008 LPSKAGIWRVECLIGEHLHAGNSTLFLVYSNK 2039
Db 1029 FPRTPGILLHCHVTDHIHAGMETTYTVLQNE 1060

RESULT 8
CERU RAT ID CERU RAT STANDARD: PRT: 1059 AA.
AC P13635; Q64719;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Ceruloplasmin precursor (EC 1.16.3.1) (Ferroxidase).
GN CP.
OS Rattus norvegicus (Rat).
OC Sukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Sprague-Dawley; TISSUE=Liver, and Lung;
RX MEDLINE=90237081; PubMed=2332446;
RA Fleming R.E., Gitlin J.D.;
RT "Primary structure of rat ceruloplasmin and analysis of
tissue-specific gene expression during development.";
RL J. Biol. Chem. 265:7701-7707(1990).
RN [2]
RP SEQUENCE OF 257-294; 571-612 AND 823-892 FROM N.A.
RC TISSUE=Liver;
RX MEDLINE=87137545; PubMed=3818625;
RA Aldred A.R., Grimes A., Schreiber G., Mercer J.F.B.;
RT "rat ceruloplasmin. Molecular cloning and gene expression in liver,
choroid plexus, yolk sac, placenta, and testis.";
RL J. Biol. Chem. 262:2875-2878(1987).
CC -I- FUNCTION: Ceruloplasmin is a blue, copper-binding (6-7 atoms per
molecule) glycoprotein found in plasma. Four possible functions
are ferroxidase activity, amine oxidase activity, copper transport
and homeostasis, and superoxide dismutase activity.
CC -I- FUNCTION: May also play a role in fetal lung development or
pulmonary antioxidant defense.
CC -I- CATALYTIC ACTIVITY: 4 Fe(2+) + 4 H(+) + O(2) = 4 Fe(3+) + 2 H(2)O.
CC -I- COFACTOR: Binds 6 Cu-ions per molecule. This protein belongs to
the multicopper oxidases which contain three distinct Cu centers


```
QY 1057 EFKKVTPLIHDRMLMDKATRLNHNMSNKTSSONMEMVQKKSGIPDPDAPDMSFF 1116
Db 733 ----- 732
QY 1117 KMLFLPESAWIORTHGKNSLNSQGPSKQLVSLGPEKSVGQNFLEKKNKVVVGKGEF 1176
Db 733 ----- 732
QY 1177 TKDVLKEMVPPSRNLEFLNLDLHNNHNNHNOBKJQEBIEKKETLQENNVLPQIHTV 1236
Db 733 ----- 732
QY 1237 TGTNFMKQLFLLSTRQNVGSDYCAVAPVLQDFPSLNSDNRKXKTAHAFSKKGBEENL 1296
Db 733 ----- 741
QY 1297 EGLGNQTKQIVKXVACTFRISNTSQQNFVQRSKRLKQFRLPLETELEKRIIVDDTS 1356
Db 742 ----- 742
QY 1357 TQWKNMHLTFTSTLQIDYNEKEKGAITQGPLSDCLTRSHSIPQANRSPDPIAKVSFP 1416
Db 743 RDWEMELHEL----- 752
QY 1417 SIRPIYLRVLFDQNSHLPAAASVKKDSGVQESSHFLQAKKNLSLAITLMTGQDR 1476
Db 753 ----- 762
QY 1477 EVGSGTSAATSVTKYKVENTVLPKPLPKTSKVELLPKVIYQKDLFPFTTSNGSPGH 1536
Db 763 ----- 762
QY 1537 LSLVEGSLQGTGAIKMNEARPGKPLRVATSSAKTPSKLLDPLANDHNGTQIPK 1596
Db 763 ----- 762
QY 1597 EEWKSEKSPKTAFKKXDTILSNACSHALAINQGNKPEIBVTWAKQGTBLRCS 1656
Db 763 ----- 762
QY 1657 QNPVLKRQREITRTTLOSOREIDYDDTISVEMKKEDFDIYDEENOSPRSFQKTRH 1716
Db 763 ----- 766
QY 1717 YFTAAVERLWDYGMSSPHVLRNRAOSGVPQKXVFOEFTDGSFTQPLNRLNEHLG 1776
Db 767 FFIGS-----KYKVVYREFDSTFREQVRRAREHLG 800
QY 1777 LLGPYIRAEVDNIMVTIERNQASRPYSFYSSLSIYBEDQCGAEPKRNFKV---PNETKT 1833
Db 801 MGLGLIHADVGAKVVKFKMARTFYSIEA-----HGVTKSSTVAPTLPGEVRT 850
QY 1834 YFKVQVQHMAPTKDEFDCKAWAYFSDVDLEKDVHSGLIGPLVC---HWTLPNAPGRQV 1890
Db 851 YIQIPERSGAGTSDSPCIWAYYSTVDRVKDLYSLGLIGLIVCRKSVKVFNPK----- 905
QY 1891 TVQEFALFFTFIBTKSWYFTENMERNCRAPCNIQMDPTFKENYRPHAINGYMDLPG 1950
Db 906 KMEFSLFLVFDENESWYLDNNTNTPDHPKDNKNDEBFIESKXKHAINGKMGFNLOG 965
QY 1951 LNPACQDQIRWYLLSMGNSNIHSIFSGHVFTVRKXKEVKMALNLYPGVFETVEMLPS 2010
Db 966 LTHVGVDENVYVWANGNEIDLEIVHFHGHGFQYKRGHSDDVDFPFGTYQILEMFPQ 1025
QY 2011 KAGIWRVECLIGEHLHAGMSTFLVYSNK 2039
Db 1026 TPGTWLLHCHVTDHAGWTTVTVLPNQ 1054
PRT: 1062 AA.
RESULT 9
ID CERU_MOUSE STANDARD; PRT: 1062 AA.
```

```
AC Q61147;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Ceruloplasmin precursor (EC 1.16.3.1) (Ferroxidase).
GN CP.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Klomp L.W.J., Farhangrazi Z.S., Choi D.W., Gitlin J.D.;
RL Submitted (MAR-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP TISSUE SPECIFICITY.
RA MEDLINE=96294736; PubMed=9690795;
RA Klomp L.W.J., Farhangrazi Z.S., Dugan L.L., Gitlin J.D.;
RL J. Clin. Invest. 98:207-215(1996).
CC -!- FUNCTION: Ceruloplasmin is a blue, copper-binding (6-7 atoms per
CC molecule) glycoprotein found in plasma. Four possible functions
CC are ferroxidase activity, amine oxidase activity, copper transport
CC and homeostasis, and superoxide dismutase activity.
CC -!- CATALYTIC ACTIVITY: 4 Fe(2+) + 4 H(+) + O(2) = 4 Fe(3+) + 2 H(2)O.
CC -!- COFACTOR: Binds 6 Cu-ions per molecule. This protein belongs to
CC the multicopper oxidases which contain three distinct Cu centers
CC known as type 1 or blue, type 2 or normal, and type 3 or coupled
CC binuclear.
CC -!- TISSUE SPECIFICITY: Expressed in many tissues, including liver,
CC eye and brain.
CC -!- SIMILARITY: Contains 3 F5/8 type A domains.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U49430; AAB07996.1; -.
DR HSSP; P00450; IKCW.
DR MGD; MGI:88476; Cp.
DR InterPro; IPR001117; Cu-oxidase.
DR InterPro; IPR002355; Cu_ox_copper_BS.
DR InterPro; IPR008972; CuproGoxin.
DR Pfam; PF00394; Cu-oxidase; 3.
DR PROSITE; PS00079; MULTICOPPER_OXIDASE1; 3.
DR PROSITE; PS00080; MULTICOPPER_OXIDASE2; 1.
DR Transprot; Ion transport; Copper transport; Oxidoreductase; Plasma;
KW Metal-binding; Copper; Repeat; Signal; Glycoprotein.
FT SIGNAL 1 19
FT CHAIN 20 1062
FT F5/8 TYPE A 1.
FT DOMAIN 20 356
FT PLASTOCYANIN-LIKE 1.
FT DOMAIN 20 199
FT PLASTOCYANIN-LIKE 2.
FT DOMAIN 208 356
FT F5/8 TYPE A 2.
FT DOMAIN 369 713
FT PLASTOCYANIN-LIKE 3.
FT DOMAIN 369 555
FT PLASTOCYANIN-LIKE 4.
FT DOMAIN 565 713
FT F5/8 TYPE A 3.
FT DOMAIN 725 1057
FT PLASTOCYANIN-LIKE 5.
FT DOMAIN 725 896
FT PLASTOCYANIN-LIKE 6.
FT DOMAIN 904 1057
FT BY SIMILARITY.
FT DISULFID 173 199
FT BY SIMILARITY.
FT DISULFID 275 356
FT BY SIMILARITY.
FT DISULFID 529 555
FT BY SIMILARITY.
FT DISULFID 632 713
FT BY SIMILARITY.
FT DISULFID 870 896
FT COPPER (TYPE 2) (BY SIMILARITY).
FT METAL 120 122
FT COPPER (TYPE 3) (BY SIMILARITY).
FT METAL 122 122
FT COPPER (TYPE 3) (BY SIMILARITY).
FT METAL 179 179
FT COPPER (TYPE 3) (BY SIMILARITY).
FT METAL 181 181
FT COPPER (TYPE 1) (BY SIMILARITY).
FT METAL 990 990
```

```

993 993 COPPER (TYPE 2) (BY SIMILARITY).
995 995 COPPER (TYPE 3) (BY SIMILARITY).
1035 1035 COPPER (TYPE 3) (BY SIMILARITY).
1036 1036 COPPER (TYPE 1) (BY SIMILARITY).
1037 1037 COPPER (TYPE 3) (BY SIMILARITY).
1041 1041 COPPER (TYPE 1) (BY SIMILARITY).
1046 1046 COPPER (TYPE 1) (BY SIMILARITY).
138 138 N-LINKED (GLCNAC. .) (POTENTIAL).
226 226 N-LINKED (GLCNAC. .) (POTENTIAL).
583 583 N-LINKED (GLCNAC. .) (POTENTIAL).
625 625 N-LINKED (GLCNAC. .) (POTENTIAL).
757 757 N-LINKED (GLCNAC. .) (POTENTIAL).
922 922 N-LINKED (GLCNAC. .) (POTENTIAL).
1062 AA; 121159 MW; F3F52ED09A239F16 CRC64;

Query Match 10.0%; Score 1243; DB 1; Length 1062;
Best Local Similarity 18.9%; Pred. No. 1.9e-56;
Matches 393; Conservative 196; Mismatches 432; Indels 1054; Gaps 30;

QY 10 FLCLLRFCF-----SATRYYLCAVELSWDMQ--SDLGELPVDARFPRVPKSPFPN 60
DB 3 FLLSTFIFLYSSLALADKHYFIGITEAVWDYASGTEKKLIISVDTQCSNPLYLQNGDR 62
QY 61 TSVVYKTLFVEFDHLENIAPRPPWGLLQPTTQAQVYDTVTITLKNMASHPVSLHAV 120
DB 63 IGRKYZKALFYDTGTSTKIDKEAMLGLFGLPVKAQVEDKVYVHLKNLASRIYTFHAR 122
QY 121 GVSYKASGABYDDQTSOREKEDDKVPFGGSHYTVQVQLKENGPMASDPLCLTYSLSH 180
DB 123 GVTYKYEYGAVPDNTDFQADDKVLPGQQVYVYVLA--NEPSPGEGDSNCVTRIYHSH 181
QY 181 VOLVQDLSGLICALLVCAEGSLAKETCTL-HKRTLLFAVDEGSKHSETKNSLM--- 236
DB 182 VDAPEDIAAGLTPGLTCKKSLYKEKEKNIDQEFVLMFVSVDENLSWLENIKTFPSE 241
QY 237 ---QDRDAASARAWPMHWNGVNRSLPGLTCHRKSVYVHWIGMTTPEVHSIFLEGH 293
DB 242 PEKVDKNEDFQESNMYSYNGTYFCSLGLSNCAADRVKWLFGMGNEVDVHSAPFHQ 301
QY 294 TFLVRNHRQASLEISITFLTAQTLMDLQGLLSCHISHOHDCMEAVKVVDSPPEEPQ 353
DB 302 ALTSRVQDTIINLPATLIDAMVAQNPQVWMLSCQNLNHLKAGLQAFQVQDC----- 356
QY 354 LINKNBERAEDDDDLTDSMDVVRDDDNSPSFIQIRSVAKKHPTWVHYTAAABEEDWD 413
DB 357 ---NKPSSKO-----NIRGKHVRH-----YYIAAEVZVWY 383
QY 414 YAP--ILVLPDDR-----SYKSQYLNNGFQIRGRKYKVRFMAYTDTF---KTREAIQH 463
DB 384 YAPSGDIFTEEBKLITAGSDPSGVFFEGATRIGGSYKMAVREYTDGSPFNKRGPDDE 443
QY 464 ESGILGPLYGEVGDILLIIFKQASRPYNIYFHGIT---DVRPLYSRKLPKGVKHLXD 519
DB 444 HLGILGPIVIAEYGDITKVTFHNKGQHUSIQMGVGSFTAEENGTYYG-----PPGASSQQA 500
QY 520 FP-ILPGEIIFYKMTVVEGPTKSDRCLTRYVSSFWANRERDLASGLIGPLLCYKESV 578
DB 501 ASHVAPKXTTYTWTVPKMGPTVADPVC-SKMYISAVDPTKOLFGLIGPMKICKGSL 560
QY 579 DQKGNQMSKRNVIILSVFEDENRSWYLTENIQRFLENPAGVQLEDFEQASNMHMSING 638
DB 561 LADGRQXVDKEFFLPTVFDENESILLDDENIRMTFHTAPQVQKEDDFQESNKMHSNMG 620
QY 639 YVFDLSQI-SVCLHEVAYWILSICAQTDLSVFPFGSYTKHKMYVEDTLTLPPFSGEIV 697
DB 621 FMYGNQWPHMCLGESITVWYLFSAAGNADVHGIFYSGNTYLCKGEEDTANLPKHSLLT 680
QY 698 FMSMENPGLWILGCHNSDFNRNGMTALLKYSSCDKNTGDDYEDSYEDISAYLLSKNAIE 757
DB 681 LMPNEDTKGTFDVECLTIDHYTGMKKQKXTVNOCOR-----QFEDFTVYL----- 724
QY 758 PRSPSONSRHSTQKQFNATTIPENDIEKTDPDWFAHRTMPKIQNVSSSDLLMLLRQSP 817

```

```

725 ----- 724
818 TPHGLSLSDIQAQAK(EIFFSDDPSPGAIDSNNSLSEWTHERPQLHSGDMVFTPEGLQLR 877
725 ----- 724
878 LNEKLGTTAAATELKCLDFKVSSTSNLIISTIPSDNLAAAGTNTSSJGPPSPMPVHYDSQD 937
725 ----- 724
938 TTLPGKKSSPLTBSGGPLS:SEBNNDSKLLIESGLMNSQESSGKGVKSVSTESGRFLKGR 997
725 -----GERT 728
998 HGPALLTKDNALFKVSTISLLKTKNTKNSATNRKXTHIDGPSLLIENSFVWQNILESUTE 1057
729 Y-----YVDA----- 733
1058 FKVTPLIHDRMLMKNATALRLNIMNKNTSSKNMEMVQOKKEGPIPPDAQNPMSEFFK 1117
734 ----- 733
1118 MLFPESARWIORTHGKNSLNSGOGSPKQI:VSLGPEKSVQGNFLSEKKNVWVGKGBFT 1177
734 ----- 733
1178 KVGGLKEMVPPSRNLFUTNLNLHENNTHNQEKKI:OBEIEKKBTLTQENVVLFOIHTVT 1237
734 ----- 733
1238 GTRKFMKNLFLSTRONVEGSDGAYAPVLQDFRSLNDSTNRTKHTAHFASKGEENLE 1297
734 -----VEVEWD--YS-- 742
1298 GLGNQTKQI:VEKYACTTILSPNTSQNFVQTSKEALKQFRLPLEETELEKRIIVDDTST 1357
743 -----SR 744
1358 QWSKXMKHLTFTLTQIDYNEKEKGAITQSPSLSDCLTRSHSIPQANRSPFLPIAKVSSPFS 1417
745 AWEKELHHL----- 753
1418 IRPIYLTRVLFODNSHLPAASRYRKSDSGVQBSSEHFLQAKKXNNLSAILILEM*GDQRE 1477
754 -----QEQNVY----- 759
1478 VGSGLTSATNSVTYKXKVENTVLPKDPKTSCKVELLPKVHIYQKDLFPTETSNGPSGHL 1537
760 ----- 759
1538 DLVEGSLLOGTGAIKWNENRPGKVPFLRVATESAKTPSKLLDPLAWNHYGTQIPKE 1597
760 ----- 759
1598 EWSQEKSPKTAFFKKOTILSNACESNHA:AAINEGONKPEIEVWTWAKQGRTERLCSQ 1657
760 ----- 759
1658 NPPVLKXHQREITRTTLQSDQEBEIDYDDTISYEMKKEDFDIYDEDENQSPRSFQKTRHY 1717
760 -----NVFLDKXEEFFI----- 770
1718 FTAABERLMDYGMSSSPHVLNRNRAQSGVQPKKVVQFQBTGSGFTQPL-YRGELNEHLG 1776
771 -----GS-KYKVVYRQDTSSSFEQVKRAEEDHEHLG 802
1777 LLOPYTRAEVEDNIMVTPANQASRPYSFYSSLSIYEDRQGAEPKKNFV---KXNETKT 1833
803 ILGPIHANVGDVKVYFKNMATRPYSIHA-----HGVKTESSTVVTPLPOEVAT 852
1834 YFWKVQHMAWPTXDEPDFCKAWAYFSDVLEKDVHESGLIGLPLVC---HTNLTLMFAHRQV 1890
853 YTWQIPERSGAGREDSACIFWAYISTVDRVKDLYSLGILGLIYVCRKSYVKVPSPK----- 907

```


QY 1891 TVQFALEFTIFDTSKSYFTENMERNCRAPCNQMDPTPKENYRPHAINGYIMDTLPG 1950
Db 908 KXMEFFFLFVFDENESYLDNKTYSSEHPEKYNKONEEFLESNKMEALNGKFGNLOG 967
QY 1951 LVMAQDQIRWYLLSMGSNENHSHSGHYFTVRKKEKQWALYNLPVGFVEVMLPS 2010
Db 968 LTMVQKDEWNYLMGMGNEDLHTVHFHGHSEFQYKRGVYSDVDFLFPFGYQILEMFPQ 1027
QY 2011 KAGLWRVSLGIERHAGMSITFLVYSNKCQTPLG 2045
Db 1028 TPGIWLHCHVTDVHAGMATYIVLPVEQETKSG 1062
RESULT 10
MFGM RAT
ID MFGM RAT STANDARD; PRT; 427 AA.
AC P70490;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Lactadherin precursor (Milk fat globule-EGF factor 8) (MFG-E8) (O-acetyl GD3 ganglioside synthase) (AGS) (MFGM).
GN MFG38 OR AGS.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP TISSUE=Brain; N.A.
RC MEDLINE=96374422; PubMed=9780713;
RX Ogura X., Nara K., Watanabe Y., Kohno K., Tai T., Sanai Y.;
RA "Cloning and expression of cDNA for O-acetylation of GD3 ganglioside."
RL Biochem. Biophys. Res. Commun. 225:932-938(1996).
CC -!- FUNCTION: May be involved in phospholipid binding. Seems to participate in the O-acetylation of GD3 ganglioside sialic acid.
CC -!- SUBCELLULAR LOCATION: Peripheral membrane protein.
CC -!- TISSUE SPECIFICITY: Spleen, lung, heart, brain and muscle.
CC -!- SIMILARITY: Contains 2 EGF-like domains.
CC -!- SIMILARITY: Contains 2 F5/8 type C domains.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation. The European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
CC
CC EMBL; D84068; BAA12210.1; -.
CC PIR; JC4915; JC4915.
CC HSP; P00740; IEDM.
CC InterPro; IPR006205; EGF like.
CC InterPro; IPR000421; FA58 C.
CC InterPro; IPR008975; GalBind like.
CC InterPro; IPR006210; IEGF.
CC Pfam; PF00754; F5_F8_type_C; 2.
CC SMART; SMO0181; EGF_2.
CC SMART; SMO0231; FA58C_2.
CC PROSITE; PS00022; EGF_1; 2.
CC PROSITE; PS01186; EGF_2; 2.
CC PROSITE; PS00026; EGF_3; 2.
CC PROSITE; PS01285; FA58C_1; 2.
CC PROSITE; PS01286; FA58C_2; 2.
CC PROSITE; PS00022; FA58C_3; 2.
CC Signal; Glycoprotein; Repeat; EGF-like domain; Milk.
FT SIGNAL 1 22 POTENTIAL.
FT CHAIN 23 427 LACTADHERIN.
FT DOMAIN 24 61 EGF-LIKE 1.
FT DOMAIN 64 108 EGF-LIKE 2.
FT

FT DOMAIN 111 267 F5/8 TYPE C 1.
FT 272 427 F5/8 TYPE C 2.
FT DISULFID 28 39 BY SIMILARITY.
FT DISULFID 33 49 BY SIMILARITY.
FT DISULFID 51 60 BY SIMILARITY.
FT DISULFID 68 79 BY SIMILARITY.
FT DISULFID 73 96 BY SIMILARITY.
FT DISULFID 98 107 BY SIMILARITY.
FT DISULFID 111 267 BY SIMILARITY.
FT DISULFID 254 258 BY SIMILARITY.
FT DISULFID 272 427 BY SIMILARITY.
FT SITE 87 89 CELL ATTACHMENT SITE (POTENTIAL).
FT CARBOHYD 61 61 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 230 230 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 280 280 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 390 390 N-LINKED (GLCNAC...) (POTENTIAL).
SQ SEQUENCE 427 AA; 47413 MW; E48C8631F3E36047 CRC64;
Query Match 5.3%; Score 663; DB 1; Length 427;
Best Local Similarity 40.3%; Pred. No. 4e-27;
Matches 142; Conservative 61; Mismatches 123; Indels 26; Gaps 6;
QY 2018 ECLGHEHLHAGKSTLFLVYSNK-----CQPLGWSAGHIRDFOITASGQY-- 2062
Db 78 KCLVTEDTQRG--DIFTEYICQCPVYSGIHCELCGCTKLGEGGAIADSQISASVYMG 135
QY 2063 ----GOWAPKLARLHYSGSINAW--STKEPFSWIKVDLLAPMIHGKITQGAQKPFSSLY 2116
Db 136 FMCLOPQWPELRLYRTGVNANWTASSYDSKPIQVDFLRKVRVSGVMTQGASRAGRAY 195
QY 2117 ISQFIIMYSLDGKKWQTYRGNSGTGLTMVFFGNVDSGGIKHNFNPITARYLHETHYS 2176
Db 196 LKTEKVAVSLDGRPEFTQDESCTGDKFPMGQDNNSLKNMNPETLEAQYIRLVPVSC 255
QY 2177 RSTLRMLMGCLNSCSMELGWSKASDAOITASSYFT--NMFA-TWSPSKARLHLOG 2233
Db 256 RCTLRFLFELGELGCGSEPLGLKNTIPDSQITASSSKTNWLRFAFGVPHGLRDNQ 315
QY 2234 RSNVRPQVNNPEKWLQVDFKTMKVTGTTQGVKSLTSMYKGLFSSQDGHQWTLF 2293
Db 316 KINAWTAQSNASKAWLQVLDLQKVKVTGIIITQGDGFHGIQYVASKVAHSDGQWTVY 375
QY 2294 FQNGKVKVFGQNDSPFPVNSLDPLLRTRYLRHPQSWHQAIALMEVLGC 2345
Db 376 EEGTSKVFQGNLDNNSHKNIPEKPFVARYVRVLPVLSWHNRITLLELLGC 427
RESULT 11
MFGM MOUSE
ID MFGM MOUSE STANDARD; PRT; 463 AA.
AC P21956; P97800;
DT 01-AUG-1991 (Rel. 19, Created)
DT 01-AUG-1991 (Rel. 19, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Lactadherin precursor (Milk fat globule-EGF factor 8) (MFG-E8) (MFGM).
GN MFG38.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 23-35.
RC TISSUE=Mammary gland;
RC MEDLINE=91046006; PubMed=2122462;
RX Stubbs J.D., Lekutis C., Singer K.L., Bui A., Yuzuki D.,
RA Srinivasan U., Parry G.;
RT "cDNA cloning of a mouse mammary epithelial cell surface protein reveals the existence of epidermal growth factor-like domains linked to factor VIII-like sequences."
RT Proc. Natl. Acad. Sci. U.S.A. 87:8417-8421(1990).
RN [2]
RP SEQUENCE OF 23-463 FROM N.A.

TISSUE=Testis;
 Besslin M.A.;
 Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
 -!- FUNCTION: May be involved in phospholipid binding. Zona pellucida-binding protein.
 -!- SUBCELLULAR LOCATION: Peripheral membrane protein.
 -!- TISSUE SPECIFICITY: Mammary epithelial cell surfaces and spermatozoan.
 -!- DEVELOPMENTAL STAGE: mRNA expression is detectable in mammary tissue from nonpregnant animals and maximal in the lactating gland.
 -!- SIMILARITY: Contains 2 EGF-like domains.
 -!- SIMILARITY: Contains 2 F5/8 type C domains.
 This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 EMBL; M38337; AAA39534.1; -;
 EMBL; Y11684; CAA72380.1; -;
 FIR; A36479; A36479.
 HSP; P00740; IEDM.
 MGD; MGI:102768; Mfeg8.
 InterPro; IPR001438; EGF I.
 InterPro; IPR006209; EGF-like.
 InterPro; IPR000421; FAS3 C.
 InterPro; IPR008979; GalBind-like.
 InterPro; IPR006210; IEGF.
 Pfam; PF00754; F5_F8 type C; 2.
 PRINTS; PR00010; EGF-BLOOD.
 SMART; SM00181; EGF; 2.
 SMART; SM00231; FAS8C; 2.
 PROSITE; PS00022; EGF_1; 2.
 PROSITE; PS01185; EGF_2; 2.
 PROSITE; PS30026; EGF_3; 2.
 PROSITE; PS01285; FAS8C_1; 2.
 PROSITE; PS01286; FAS8C_2; 2.
 PROSITE; PS50022; FAS8C_3; 2.
 Signal; Glycoprotein; Repeat; EGF-like domain; Milk.
 SIGNAL 1 22
 CHAIN 23 463
 DOMAIN 24 61
 DOMAIN 64 108
 DOMAIN 148 303
 DOMAIN 308 463
 SITE 87 89
 SITE 28 39
 DISULFID 33 49
 DISULFID 51 60
 DISULFID 68 79
 DISULFID 73 96
 DISULFID 98 107
 DISULFID 149 303
 DISULFID 290 294
 DISULFID 308 463
 CARBOHYD 61 61
 CARBOHYD 266 266
 CARBOHYD 316 316
 CARBOHYD 426 426
 CONFLICT 30 35
 CONFLICT 35 35
 CONFLICT 110 147
 CONFLICT 168 168
 CONFLICT 196 196
 CONFLICT 309 309
 CONFLICT 395 395
 CONFLICT 463 AA; 51465 NW; D78B6C6E9FBA724D CRC64;
 SEQUENCE

Query Match 5.3%; Score 657; DB 1; Length 463;
 Best Local Similarity 42.2%; Pred. No. 9.3e-27;
 Matches 135; Conservative 63; Mismatches 110; Indels 12; Gaps 5;
 QY 2037 SNKCTPLGMSGHIRDFQITAGQY-----GQWAPKLARLHYSGSINAW--STKEPFS 2088
 DB 145 ASRCSTQLGMEGGAIDSQISASVYVFMGLQRMGPELARLYRTGIVNAWNASNYDSKP 204
 QY 2089 WIKVDLLAPMIHGIKTQGRARQFSSLYISQFIIVYSLDGKKWOTYRGNSTGTLMVFFGN 2148
 DB 205 WIQVNLLEKRVSGWMTQSGARAGRYLKTFFKAYSIDGRKKPFIODESGD-KEFLGN 263
 QY 2149 VDSGIGKHTFNPIIARIYRLHPTYSIRSLRMELMGDLNCSPLGNESKAISDAQ 2208
 DB 264 LDNLSKVNENPLEAQYRLYFVSHRGCTLRFELLGCELHGLPLGKNTIPDSQ 323
 QY 2209 ITASSYFT--NMFA-TWSPSKARLHLOGRNAPVQNNPKWLOVDFOKTMKVTGVITQ 2265
 DB 324 MSASSYKTNLRAFQWYPLHGLRDLNQGKINAWTAQNSAKEWLQVDLGTQROVTGIITQ 383
 QY 2266 GVKSLTSMYKELFSLSSQDCHOWTLFFQNGKVKVFOGNODSTFPVYVNSLDPLLTRYL 2325
 DB 384 GARDPGR:IOYVESYKVAHSDDGQVMTVEQSSKVFQGNLDNNSHKNIKPEPMARYV 443
 QY 2326 RIHQSVVHQIALEMEVLGC 2345
 DB 444 RVLPYSWNRITLRLLELGC 463
 RESULT 12
 MFGM FIG STANDARD; PRT; 403 AA.
 ID MFGM FIG
 AC 279385;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 15-MAR-2004 (Rel. 43, Last annotation update)
 DE Lactadherin (Milk fat globule-EGF factor 8) (MFG-8) (MFGM) (Spertm surface protein Sp47) (PP47).
 GN MFG8.
 OS Sus scrofa (Pig).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
 CC NCBI_TaxID=9823;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Testis;
 RA Besslin M.A.;
 RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
 CC -!- FUNCTION: May be involved in phospholipid binding. Zona pellucida-binding protein.
 CC -!- SUBCELLULAR LOCATION: Peripheral membrane protein.
 CC -!- TISSUE SPECIFICITY: Mammary epithelial cell surfaces and spermatozoan.
 CC -!- SIMILARITY: Contains 2 EGF-like domains.
 CC -!- SIMILARITY: Contains 2 F5/8 type C domains.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC EMBL; Y11683; CAA72379.1; -;
 CC FIR; T11743; T11743.
 CC HSP; P00740; IEDM.
 CC InterPro; IPR006209; EGF-like.
 CC InterPro; IPR000421; FAS3 C.
 CC InterPro; IPR008979; GalBind-like.
 CC InterPro; IPR006210; IEGF.
 CC Pfam; PF00008; EGF; 2.

```
DR Pfam: PF00754; F5_P8_type_C; 2.
DR SMART; SM00181; EGF; 2.
DR RA SMART; SM00231; FAS8C; 2.
DR DR PROSITE; PS00022; EGF_1; 2.
DR DR PROSITE; PS01186; EGF_2; 2.
DR DR PROSITE; PS00026; EGF_3; 2.
DR DR PROSITE; PS01285; FAS8C_1; 2.
DR DR PROSITE; PS01285; FAS8C_2; 2.
DR DR PROSITE; PS01286; FAS8C_3; 2.
DR DR PROSITE; PS00022; FAS8C_4; 2.
KW Glycoprotein; Repeat; EGF-like domain.
FT DOMAIN 2 41 EGF-LIKE 1.
FT DOMAIN 44 88 EGF-LIKE 2.
FT DOMAIN 91 247 F5/8 TYPE C 1.
FT DOMAIN 252 409 F5/8 TYPE C 2.
FT SITE 67 69 CELL ATTACHMENT SITE (POTENTIAL).
FT SITE 6 17 BY SIMILARITY.
FT DISULFID 11 29 BY SIMILARITY.
FT DISULFID 31 40 BY SIMILARITY.
FT DISULFID 91 247 BY SIMILARITY.
FT DISULFID 234 238 BY SIMILARITY.
FT DISULFID 252 409 BY SIMILARITY.
FT CARBOHYD 41 41 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 372 372 N-LINKED (GLCNAC...) (POTENTIAL).
SQ SEQUENCE 409 AA; 45725 MW; BOC07AF8002927A CRC64;

Query Match 5.2%; Score 650; DB 1; Length 409;
Best Local Similarity 39.3%; Pred. No. 1.8e-26;
Matches 139; Conservative 62; Mismatches 125; Indels 28; Gaps 6;

QY 2018 ECLIGEHLAGMSLFVYSNK-----CQPLGMSAGHIRDFOITASGOY-- 2062
D6 58 ECEVIDAHRG--DVPEYICKPHGYTGHCILICNAPLGMETGATDFQISASSHLG 115
QY 2063 ----GOWAPKLARIFYGGSINAW--STKEPSWTKVDLLAPMIHGHKTQARQKFSLSY 2116
D6 116 FMGLQRAWPELRLHRAIGIVNAWTASNDYRNPQVNLRLRRVYGVVTOGASAGSAEY 175
QY 2117 ISQFIIVSLDCKKQWOTRGSTGTLMVFFGNVDSGSKHNIENPPIIARVIRLHPHYS 2176
D6 176 MKTFKVAYSTDGRKFQFIQGAESGDKIFMGLNDSGLKVLNLFVPELVQVRLVPIICH 235
QY 2177 IRSTLRLMELMGCDLNSCMLPMSKATSDAQITASSYFTN---MFAWSPSKARLHLQG 2233
D6 236 RGCTLRPELLGCELSCGAELGLKNDTTPNKGITASSFYRTWGLSAFWSYFYARLDNQ 295
QY 2234 RSNARPOVNNKQWLVDFQKMTKVGTGTQGVKSLTSTMYKVELTSSQDGHOWTLF 2293
D6 296 KFNAWTAQSNASEWLQIDLGSQREVGTGITQGDQDFGHIQVAAKYAVYSDDGVSWTET 355
QY 2294 PQNGKV--KVFQGNDSFTPVVNSLDPPLLTFLYLRHPQSWHQAIRMEVLGC 2345
D6 356 RDQALGKIFPNLNDNNSHKQMPETPLFRFVRLPVANHRITLAVELGCG 409

RESULT 13
ID MFGM_BOVIN STANDARD; PRT; 427 AA.
AC Q95114; F79344; Q27959;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Lactadherin precursor (Milk fat globule-EGF factor 8) (MFG-E8)
DE (MG57/53) (PAS-6/PAS-7 glycoprotein) (MFGM) (Sperm surface protein
DE SP47) (BP47) (Components 15/16).
GN MFG88.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A., PARTIAL SEQUENCE, AND CARBOHYDRATE-LINKAGE SITES.
RQ STRAIN=Holstein; TISSUE=Mammary gland;
```

```
RX Pfam: PF008954; PubMed=8856064;
RA Hvaregaard J., Andersen M.H., Berglund L., Rasmussen J.T.,
RA Petersen T.E.;
RT "Characterization of glycoprotein PAS-6/7 from membranes of bovine
RT milk fat globules.";
RL Eur. J. Biochem. 240:628-635(1996).
RN [2]
RP SEQUENCE OF 18-427 FROM N.A.
RC TISSUE=Mammary gland;
RX MEDLINE=96125736; PubMed=8541316;
RA Aoki N., Kishi M., Taniguchi Y., Adachi T., Nakamura R.,
RA Matsuda T.;
RT "Molecular cloning of glycoprotein antigens MGPS7/53 recognized by
RT monoclonal antibodies raised against bovine milk fat globule
RT membrane.";
RL Biochim. Biophys. Acta 1245:385-391(1995).
RN [3]
RP SEQUENCE OF 19-427 FROM N.A.
RC TISSUE=Testis;
RA Esslin M.A.;
RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE OF 140-146; 174-187; 233-246 AND 422-427.
RC TISSUE=Milk;
RX MEDLINE=93250576; PubMed=8485470;
RA Mather I.H., Banghart L.R., Lane W.S.;
RT "The major fat-globule membrane proteins, bovine components 15/16 and
RT guinea-pig GP 55, are homologous to MGF-E8, a murine glycoprotein
RT containing epidermal growth factor-like and factor V/VIII-like
RT sequences.";
RL Biochem. Mol. Biol. Int. 29:545-554(1993).
CC -!- FUNCTION: Probably associates with phospholipids on the surface of
CC mammary epithelial cells and milk fat globules. Zona pellucida-
CC binding protein.
CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Name=Long;
CC IsoId=Q95114-1; Sequence=Displayed;
CC Name=Short;
CC IsoId=Q95114-2; Sequence=VSP_001398;
CC -!- TISSUE SPECIFICITY: Milk and spermatozoan.
CC -!- PTM: THE 2 O-LINKED GLYCANS CONSIST OF GAL, GLCNAC AND FUC, WITH
CC PROBABLY FUC AS REDUCING TERMINAL SUGAR.
CC -!- SIMILARITY: Contains 2 EGF-like domains.
CC -!- SIMILARITY: Contains 2 F5/8 type C domains.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
```



```
DR InterPro; IPR006210; EGF.
DR Pfam; PF00008; EGF; 1.
DR Pfam; PF00754; F5_F8_type_C; 2.
DR PRINTS; PR00010; EGFBL00D.
DR SMART; SM00181; EGF; 1.
DR SMART; SM00231; FAS8C; 2.
DR PROSITE; PS00022; EGF_1; 1.
DR PROSITE; PS01186; EGF_2; 1.
DR PROSITE; PS00026; EGF_3; 1.
DR PROSITE; PS01285; FAS8C_1; 2.
DR PROSITE; PS01286; FAS8C_2; 2.
DR PROSITE; PS00022; FAS8C_3; 2.
KW Signal; Glycoprotein; Milk; Repeat; EGF-like domain; Amyloid.
FT SIGNAL 1 23
FT CHAIN 24 387
FT CHAIN 202 387
FT CHAIN 268 317
FT CHAIN 24 67
FT DOMAIN 70 225
FT DOMAIN 230 387
FT SITE 46 48
FT SITE 27 38
FT DISULFID 32 55
FT DISULFID 57 66
FT DISULFID 70 225
FT DISULFID 212 216
FT DISULFID 230 387
FT CARBOHYD 238 238
FT CARBOHYD 325 325
FT CARBOHYD 329 329
FT CARBOHYD 350 350
FT SEQUENCE 387 AA; 43123 MW; 286571DEC83782D CRC64;
Query Match 4.7%; Score 588; DB 1; Length 387;
Best Local Similarity 37.3%; Pred. No. 2.6e-23;
Matches 132; Conservative 69; Mismatches 125; Indels 28; Gaps 9;
QY 2006 EMLPSKAGIWRVCELGHEHAGMSTFLVYVSKCOTPLGMSGHIRDFQITASG---- 2060
DB 48 DFFPS-----YTCTLKG---YAGNCG-----STKCVFPLMGNGNIAQSIASSVRVTF 95
QY 2061 -QYGWAPKIALHYSGSINAW--STKEPFSIKVDLLAPMIHGIKTQARQKSSLY 2117
DB 96 LCLQHWVELARLNACGMVNAWTPPSNDNDPIQVNLRRXVWTGVVQGSRLASHEYL 155
QY 2118 SDFILYLDGKQWQYRGNSTGTLMVFGNVDSGKIKNINPPIARIYILHPTHSI 2177
DB 156 KAFKVAVSLNCHFEFD-FIHDVNNKKHKEFVGNWNNKNAVHVNLFETPYEAQYVLYPTSCHT 214
QY 2178 RSTLRELMLGCDLNSCMPLGWSKXISDAQTASSYF---TNMPATWSPSKARHLQOG 2233
DB 215 ACTLRFELLCGLCNGANPLGLKNNIPDKQITASSSYKTWGLHLP-SNPFSYARLDKQG 273
QY 2234 RSNAPRPQVNNKREWLOQDFQKTKMTKVTGTQGVKSLTSMVYKFLISSODGHQWTLF 2293
DB 274 NFNWAVAGSYGNDQWLQVDLGSKEVTGTLTGARNFSGVQFVASKVAYNSDANWTEY 333
QY 2294 F--QNGKVKVQGNQDSFTFVNSLDDPLLTRVLRHPQSWHQAIRMEVLGC 2345
DB 334 QDPRTGSGKIFFGNWNBHSHKKNLFTPIIARYVRLPVAWENRIARLELLGC 387
RESULT 15
ID NRP2_HUMAN STANDARD; PRT; 931 AA.
AC O60452; O14820; O14821;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Neuropilin-2 precursor (Vascular endothelial cell growth factor 165
DE receptor 2).
GN NRP2 OR VEGF165R2.
OS Homo sapiens (Human).
```

```
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OC NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORMS A0 AND A17).
RX MEDLINE=97470888; PubMed=9331348;
RA Chen H., Chedotal A., He Z.-G., Goodman C.S., Tessier-Lavigne M.;
RT "Neuropilin-2, a novel member of the neuropilin family, is a high
RT affinity receptor for the semaphorins Sema E and Sema IV but not Sema
RT III.";
RL Neuron 19:547-559(1997).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORM A22).
RX TISSUE=Breast;
RA Soker S., Takashima S., Miao H.-Q., Neufeld G., Klagsbrun M.;
RT "Neuropilin-1 is expressed by endothelial and tumor cells as an
RT isoform-specific receptor for vascular endothelial growth factor.";
RL Cell 92:735-745(1998).
RN [3]
RP CHARACTERIZATION.
RX MEDLINE=20309748; PubMed=10748121;
RA Gluzman-Poltorak Z., Cohen T., Herzog Y., Neufeld G.;
RT "Neuropilin-2 and neuropilin-1 are receptors for the 165-amino acid
RT form of vascular endothelial growth factor (VEGF) and of placenta
RT growth factor-2, but only neuropilin-2 functions as a receptor for
RT the 145-amino acid form of VEGF.";
RL J. Biol. Chem. 275:18040-18045(2000).
CC -!- FUNCTION: High affinity receptor for semaphorins 3C, 3F, VEGF-165
CC and VEGF-145 isoforms of VEGF, and the PLGF-2 isoform of PGF.
CC -!- SUBUNIT: Neuropilin-2 probably forms a heteromeric complex with
CC neuropilin-1 in order to be a functional semaphorin 3C receptor.
CC -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC -!- ALTERNATIVE PRODUCTS:
CC Event-Alternative splicing; Named isoforms=3;
CC Name=A22; IsoId=O60462-1; Sequence=Displayed;
CC Name=A0; IsoId=O60462-2; Sequence=VSP_004342;
CC Name=A17; IsoId=O60462-3; Sequence=VSP_004341;
CC -!- SIMILARITY: Belongs to the neuropilin family.
CC -!- SIMILARITY: Contains 2 CUB domains.
CC -!- SIMILARITY: Contains 2 F5/8 type C domains.
CC -!- SIMILARITY: Contains 1 MAM domain.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AF022859; AAC51788.1; -
CC EMBL; AF022860; AAC51789.1; -
CC EMBL; AF016098; AAC12922.1; -
CC HSPB; P12259; 1CZT.
CC Genew; HGNC:8005; NRP2.
CC NIM; 602070; -
CC GO; GO:0005624; C:membrane fraction; TAS.
CC GO; GO:0004872; F:receptor activity; TAS.
CC GO; GO:0005021; F:vascular endothelial growth factor receptor. . .; TAS.
CC GO; GO:0007411; P:axon guidance; TAS.
CC InterPro; IPR000859; CUB.
CC InterPro; IPR000421; FAS8_C.
CC InterPro; IPR008979; Gal_bind-like.
CC InterPro; IPR000998; MAM_domain.
CC Pfam; PF00431; CUB; 2.
CC Pfam; PF00754; F5_F8_type_C; 2.
CC Pfam; PF00629; MAM; 1.
CC PRINTS; PR00020; MAMDOMAIN.
CC SMART; SM00042; CUB; 2.
```

QY 2257 MKVTGVTTQG-----VKSLTSMYKVEPLISSQDGHQWTLFQNGKV---KVFQGNQD 2307
Db 496 KTVKGVIIQARGGDSITAVEARAFVRKFKVSYSLNGKWE-YIQDPRTOQPKLFEQNMH 554
QY 2308 SFTPVNSLDPPPLTLRLRIHPQSVWHQ-IALRMEVLGCEAQD 2349
Db 555 YDTPDIRFD-PIPAQYRVYPERWSPAGIGMRLEVLGCDWTD 596

Search completed: April 13, 2004, 14:05:28
Job time : 38 secs

DR SMART; SM00231; FA58C; 2.
DR SMART; SM0137; MAM; 1.
DR PROSITE; PS01180; CUB; 2.
DR PROSITE; PS01285; FA58C; 1; 2.
DR PROSITE; PS01286; FA58C; 2; 2.
DR PROSITE; PS01286; FA58C; 3; 2.
DR PROSITE; PS00022; FA58C; 3; 2.
DR PROSITE; PS50060; MAM; 2.
KW Transmembrane, Glycoprotein; Neutrone; Signal; Repeat; Receptor;
KW Alternative splicing.
FT SIGNAL 1 20
FT CHAIN 21 931
FT DOMAIN 21 864
FT TRANSMEM 865 889
FT DOMAIN 890 931
FT DOMAIN 28 142
FT DOMAIN 149 267
FT DOMAIN 277 427
FT DOMAIN 434 592
FT DOMAIN 642 802
FT DOMAIN 671 674
FT DISULFID 28 55
FT DISULFID 83 105
FT DISULFID 149 175
FT DISULFID 208 230
FT DISULFID 277 427
FT DISULFID 434 592
FT CARBOHYD 152 152
FT CARBOHYD 157 157
FT CARBOHYD 629 629
FT CARBOHYD 839 839
FT VARSPLIC 809 813
FT FTID-VSP 004341.
FT VARSPLIC 809 830
FT CONFLICT 602 602
FT SEQUENCE 931 AA; 104830 MW; 270CBAB69A0A797C CRC64;
Query Match 3.8%; Score 469.5; DB 1; Length 931;
Best Local Similarity 28.8%; Pred No. 1.3e-16;
Matches 166; Conservative 87; Mismatches 195; Indels 145; Gaps 32;
QY 1847 DEEDCK-AWAYFSDVDLEK----DVHSLGIGPLLVCHNTYL-----NPAHGRQVTVQBEFA 1896
Db 79 EKHDCVDFEIRDGSEADLLGKHGNTAPPTIISGSMYIKFTSDYARQCA--GFS 136
QY 1897 LFTIPDETYSWYFTENMERNCRAPCIQMEDPTEKENYRHALNGYIMDTLPLGYMAQD 1956
Db 137 LRYEIRF-KTGS-----EDCSKNFTSP-KGTIESPGFPEKIP----- 170
QY 1957 QRIRWLLSGMSENTHSIHPSGHVFTVRKKEEYKMAKLYLPGVFTVEMLPKAG--- 2013
Db 171 -----HNLDCTFTILAKFKWEILQLF---IFD-LSHDPLQVGECD 207
QY 2014 -----IW-----RVECLIGEH-----LHAGNST-----LFLY 2035
Db 208 CKYDWDLDWDGPHVGLPGICKYCTKTPSELRSSTGLSTLFTHTDMAVAKGFSARYLY 267
QY 2036 YSN-----KCTPLGMASGHIRDQITASGOY--GOWAPKLRLHYGSGINAW-----ST 2083
Db 268 HQELENFQCNVPLGNESGRIANEQISASTYSQGRWTPQOSRLH--GDDNGWTFNLDN 325
QY 2084 KEPTSWIKVDLLAEWIIHGKITQGA--RQKPSLIYSQFIIVYSLDGKKWOTYRGNSTGT 2141
Db 326 K3---YLOVDLRFLLMTLTAITATQGAISRETQNGYVYKYLEVSTNGEDMWYRHKNNH- 381
QY 2142 LMVFGNVDSGKIKHINFPPIIARYELHPTHYSIRSTRMELMGCDLNS--CSMPFGM 2199
Db 382 -KVFOANNDATEVVLNKLHAPLLTRFVRIRPQTWHSGIALRLFLGCRVTDAPCSNMLGM 440
QY 2200 ESKAISDAQITASSYFTNMEATWSPSKARLHLCQGSNNAW---RPQVNNPKEMLOVDFOKT 2256
Db 441 LSGLIADSQISASS--TQFY-LWSPSAARL-VSSRS-GWFPRIPOAQPGEEWLVLDLGTGP 495

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 13, 2004, 14:01:27 ; Search time 68.5 Seconds

(without alignments)
10828.953 Million cell updates/sec

Title: NP000123-328-355-581

Perfect score: 12415

Sequence: 1 MQELSTCFCLLRFCFSA.....WVHQIALRMEVLGCEAQQDIY 2351

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SPTREMBL 25:*

- 1: sp_archaea:*
- 2: sp_bacteria:*
- 3: sp_fungi:*
- 4: sp_human:*
- 5: sp_invertebrate:*
- 6: sp_mammal:*
- 7: sp_mhc:*
- 8: sp_organelle:*
- 9: sp_phase:*
- 10: sp_plant:*
- 11: sp_rodent:*
- 12: sp_virus:*
- 13: sp_vertebrate:*
- 14: sp_unclassified:*
- 15: sp_virus:*
- 16: sp_bacteriaph:*
- 17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	9474	76.3	2343	6 O18806	O18806 canis fam1
2	9428	75.9	2343	6 O62730	O62730 canis fam1
3	6276.5	50.6	2258	11 Q7TN96	Q7TN96 rattus norv
4	3034.5	24.4	1639	13 Q804W6	Q804W6 fugu rubrip
5	2823.5	22.7	2119	13 Q90X47	Q90X47 brachydanio
6	2771.5	22.3	2183	11 Q88783	Q88783 mus musculu
7	2685	21.6	1802	13 Q804W5	Q804W5 fugu rubrip
8	2511.5	20.2	1377	13 Q804X3	Q804X3 gallus gall
9	2343	18.9	1460	13 Q7SZN0	Q7SZN0 pseudonaja
10	1999.5	16.1	2102	11 Q7TPK2	Q7TPK2 rattus norv
11	1595.5	12.9	355	11 Q8BQ43	Q8BQ43 mus musculu
12	1433	11.7	745	13 Q804X4	Q804X4 gallus gall
13	1390	11.2	1156	11 Q80Y80	Q80Y80 mus musculu
14	1389	11.2	1157	11 Q920Z4	Q920Z4 mus musculu
15	1375	11.1	1157	11 Q920H8	Q920H8 rattus norv
16	1345	10.8	1156	4 Q9BQS7	Q9BQS7 homo sapien

Q9C058 homo sapien
Q7ZU12 brachydanio
Q75180 homo sapien
Q91197 rattus norv
Q9XT27 ovis aries
Q8C4S2 mus musculu
Q14286 homo sapien
Q90ZT2 brachydanio
Q8AYE0 brachydanio
Q81191 chlamydomon
Q8BV37 mus musculu
Q9TLX9 mus musculu
Q8N610 homo sapien
Q43854 homo sapien
Q9WTS3 mus musculu
Q8C4U8 mus musculu
Q8CBF7 mus musculu
Q35474 mus musculu
Q7Z3D2 homo sapien
Q77718 equus caball
Q9BTL9 homo sapien
Q9H2E4 homo sapien
Q9H2D5 homo sapien
Q9H2D4 homo sapien
Q9H2E3 homo sapien
Q7ZC39 homo sapien
Q8C8X0 mus musculu
Q8QZY7 mus musculu
Q9GX38 rattus norv

ALIGNMENTS

RESULT 1

O18806 ID O18806 PRELIMINARY; PRT; 2343 AA.
AC O18806;
DT 01-JAN-1998 (TREMBLrel. 05, Created)
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE Factor VIII.
GN F8.
OS Canis familiaris (Dog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
OX NCBI_taxID=9615;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RA Cameron C., Notley C., Hoyle S., McGlynn L., Hough C., Kamisue S.,
RA Giles A., Lillicrap D.;
RT "The canine factor VIII cDNA and 5' flanking sequence."
RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: CONTAINS 2 F5/8 TYPE C DOMAINS.
DR EMBL; AF016234; AAB87412.1; -.
DR HSSP; P00451; 1CRF.
DR GO; GO:0005507; F:copper ion binding; IEA.
DR GO; GO:0007155; P:cell adhesion; IEA.
DR InterPro; IPR001117; Cu-oxidase.
DR InterPro; IPR008972; Cupredoxin.
DR InterPro; IPR000421; FA58 C.
DR InterPro; IPR008979; Gal_bind_like.
DR Pfam; PF00754; Cu-oxidase; 3.
DR Pfam; PF00754; F5_F8 type C; 2.
DR SMART; SM00231; FA58C_2; 2.
DR PROSITE; PS01285; FA58C_1; 2.
DR PROSITE; PS01286; FA58C_2; 2.
DR PROSITE; PS00022; FA58C_3; 2.
DR PROSITE; PS00079; MULTICOPPER OXIDASE1; 3.
SQ SEQUENCE 2343 AA; 265829 MW; A854FAE571C3B399 CRC64;

Query Match

76.3%; Score 9474; DB 6; Length 2343;


```
QY 2153 GIKHNIENPPIIAYIRLHPHYSIRSLRMLMGLCDLNSCSMPLGMSKASISDAQITAS 2212
DB 2145 GIKHNIENPPIIAYIRLHPHYSIRSLRMLMGLCDLNSCSMPLGMSKASISDAQITAS 2204
QY 2213 SYFTNMEATWSPSKARLHLOGSNAPVPOVNPXEWLQVDFQKTMKVTVGTTOGVKSLLT 2272
DB 2205 SYLSMIAWSPSQRHLHQRTNAWRPOANPKWLVQDFRKMVKVGTGTTQGVKSLLI 2264
QY 2273 SMYKFEFLISSQGHOWTLFFQNGKVKVFGQNDSTFPVVSNSIDPPLRLYLRHPQSW 2332
DB 2265 SMYKFEFLISSQGHENWTLFLQNGKVKVFGQNRDSTFPVNRLEPPLVARYVRLHPQSW 2324
QY 2333 VHQAALMEVLGCEAQ 2348
DB 2325 AHHIALREVLGCDTQ 2340

RESULT 2
O62730 PRELIMINARY; PRT; 2343 AA.
AC O62730;
DT 01-AUG-1998 (TREMELrel. 07, Created)
DT 01-AUG-1998 (TREMELrel. 07, Last sequence update)
DT 01-OCT-2003 (TREMELrel. 25, Last annotation update)
DE Factor: VIII.
OS Canis familiaris (Dog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
OX NCBI_TaxID=9615;
XN (1)
XP SEQUENCE FROM N.A.
RC TISSUE=Kidney, and Spleen;
RA Gordy P.W., Bowen R.A.;
RT "Characterization of the canine factor VIII cDNA.";
RL Submitted (FEB-1998) to the EMBL/GenBank/DBJ Databases.
CC -1- SIMILARITY: CONTAINS 2 F5/8 TYPE C DOMAINS.
DR BMBI; AF043489; RAC05384.1; -.
DR HSSP; P00451; 1CFG.
DR GO; GO:0005507; F: copper ion binding; IEA.
DR GO; GO:0017155; P: cell adhesion; IEA.
DR InterPro; IPR001117; Cu-oxidase.
DR InterPro; IPR008972; Cupredoxin.
DR InterPro; IPR000421; FA58_C.
DR InterPro; IPR008979; Gal_bind_like.
DR Pfam; PF00394; Cu-oxidase; 3.
DR Pfam; PF00754; F5_F8_type_C; 2.
DR SMART; SM00231; FA58C_2.
DR PROSITE; PS01285; FA58C_1; 2.
DR PROSITE; PS01286; FA58C_2; 2.
DR PROSITE; PS00022; FA58C_3; 2.
DR PROSITE; PS00079; MULTICOPPER_OXIDASE1; 3.
SQ SEQUENCE 2343 AA; 265613 MW; F612D744ADAADD99 CSC64;

Query Match 75.9%; Score 9428; DB 6; Length 2343;
Best Local Similarity 76.7%; Pred. No. 0;
Matches 1806; Conservative 200; Mismatches 326; Indels 24; Gaps 12;

QY 1 MQELSTCPFLCLLRFCFSATRYYLGAVELSDYMQSD-LGELPYDARPPPVKSPFF 59
DB 1 MQVELYTCFLCLLPFLSLSATRYYLGAVELSDYMQSDLLSALHADTSFSSRPVGLPL 60
QY 60 NTSVYVKTTFVFTDHLFIAPRPWMLGLGPTQASVYDVTWITLKNWASHPVSLHA 119
DB 61 TTSVYVKTTFVFTDHLFIAPRPWMLGLGPTQAEVIDIVVILKNWASHPVSLHA 120
QY 120 VGSYVXASGAEDDQTSREKEDDRVFPFGSHYVQVLKENGPNWASDPLCLITYSL 179
DB 121 VGSYVXASGAEDDQTSREKEDDNVIFGESHYVQVLKENGPNWASDPLCLITYSL 180
QY 180 HVDLVKDLNGLIGALLVCRFGSLAKSKTQTLHLKFIILLFAVFDGKGWHSKTSNLMQDR 239
DB 181 HVDLVKDLNGLIGALLVCRFGSLAKSKTQTLQEVLLFAVFDGKGWHSKTSNLTQ-- 238
```

```
QY 240 DAASARAWPKMHTVNGYVNRSLPGLICCHRSKYVHWIGMGTTPPEVHSIFLEGHTFLVRN 299
DB 239 ----AEAQHELHITNGYVNRSLPGLVCHRSKYVHWIGMGTTPPEVHSIFLEGHTFLVGN 294
QY 300 HRQASLEISITPFLTAQTLMDLGOFLLSCHLSSHQHDGMEAYVKVDSCEEPQLMKNN 359
DB 295 HRQASLEISITPFLTAQTLMDLGOFLLFCHPSHQHDGMEAYVKVDSCEEPQLMKNN 354
QY 360 EEAEDYDDITDSEMDVVRFDNNSPSFIQIRSVAKKHPTKWHYIAAEBEDWDYAPLVL 419
DB 355 ED-KOYDDGLYGSMDVVSFDDSSSPFIQIRSVAKKHPTKWHYIAAEBEDWDYASGP 413
QY 420 APDDRSYKSVLNNQPQIRIGRKYKVFMAYTDETFKTREAIQHSIGILPPLYGEGDT 479
DB 414 TPNDRSHKNLYLNNQPQIRIGKCKKVFMAYTDETFKTREAIQHSIGILPPLYGEGDT 473
QY 480 LLIIFKNOQASRPNIYPHGIDVREPLYSRRLPKVKHLKOPPLPGLPGIEKYKMTVTVEDG 539
DB 474 LLIIFKNOQASRPNIYPHGINYVTPLHTGRLPKVKHLKOPPLPGLPGIEKYKMTVTVEDG 533
QY 540 PTKSDPRCLTRYSGSVNMRDLASGLIGPLLI CYKESVDQKGNQIMSKDNVILFSYVD 599
DB 534 PTKSDPRCLTRYSGSVNMRDLASGLIGPLLI CYKESVDQKGNQIMSKDNVILFSYVD 593
QY 600 ENRSWYLTENIOTRELPNPAGVQLEDPPFQASINMHSINGYVDFSLQLSVC-HEVAYWYL 659
DB 594 ENRSWYLTEDMQRELPNADVVQPHDPFQLSNIMHSINGYVDFSLQLSVC-HEVAYWYL 653
QY 660 SIGAQTDLSVFFSGYTFKHWYEDTLTPFSGSETVFMENPGLWILGCHNSDFNR 719
DB 654 SVGAQTDLSVFFSGYTFKHWYEDTLTPFSGSETVFMENPGLWILGCHNSDFNR 713
QY 720 GMTALLKVSQCDKGTGYEDSYEDI SAYLLSKNNAIEFRSFSQNSRHPSTROKQFNATT 779
DB 714 GMTALLKVSQCNRIIDDYEYEDIPPLANENNVIKERSFSQNSRHPSTROKQFNATT 773
QY 780 IPENDIEKTDPMFAHRTPMKIQNVSSDILLMLRQSTPHGLSLDSLOEAKYETFSDDP 839
DB 774 IPENDIEKIDLQSGERTQILKAQSVSSDILLMLLQGNPTPRGLFSLDREATDR--ADGH 831
QY 840 SPGALDSNNLSSENTHFRPOLHSGDMVTPESGLOLRNEKLGTAATAELKDLKDFVSS 899
DB 832 SRGAIERNKGPPEVASLRPELHSHSDBRFTPPPELQRLNENLGNVTVELKKDLKISS 891
QY 900 TSNLLIS--TIPSDNLAAGTDNTSIGPPSPVHYDSDQDITLFGKSSPLTESGGPUL 957
DB 892 SSDSLMTSPTIPSDKLAATEKTGSLGPPENMSVHENGHLGTVFGNNSHLLIQSVPLEL 951
QY 958 SEENDSKLLESGLMNSQESSWGKQVSVSTESGRLPKGKAHGFALLTKDNALPKYSISLL 1017
DB 952 SEENDSKLLEAPLMNIQESSLRENVLMSNRLFKEEIRGPASLIKDNALPKYNISSV 1011
QY 1018 KTKNTSNRSATNREKTHIDGSPILLIENSPSVWONI-LESDETFKVTPLIHDRMLMDKNAT 1076
DB 1012 KTNRAPVNTTKRUTKVAITPILLIENSTSVWDIMLERTFEKVTSLIHNETFMDRNTT 1071
QY 1077 ALRLHMSNKTSSKNMVMVQKKEGPIPPDAQNPDMSPFKMLFPEASARWQTHGKNS 1136
DB 1072 ALGLNHSNKTLLSKNVEVAHQKEDVPVLAENPDLSKKIPFLPD--WI-KTHGKNS 1127
QY 1137 LNSGQSPSKQLVSIQPEKSVGEQNFSLSEKNVYVCKGGETKDVGLKEMVFPSSRNLFLT 1196
DB 1128 LSSEQSPSKQLVSIQPEKSVGEQNFSLSEKNVYVCKGGETKDVGLKEMVFPSSRNLFLT 1185
QY 1197 NLDNLHENHNTQEKKEKIEIEKTELIQENVVLPOIHTVTGKKNMKMLFLISTRQNYE 1256
DB 1186 NLANQVENQTYQEKKSLEIERKEKLTQENVALPOAHTWIGTKNFKNLFLISTQNYA 1245
QY 1257 GSYDGAAPVLQDFRSLNDSNTRTKHTAFSKKGEENLEGNGNQTKQIVKAYACTRI 1316
DB 1246 GLEEQYTPILQDTRSLNDSPHSEGHMFANFSKIREANLEGLNGNQTNQWVERFPSTRM 1305
```

```
QY 1317 SPNTSQNFVTCRKBALQKQSLPLEETELEKRIIVDDTSTQWKNMHHLPSTLTQIDY 1376
Db 1306 SSNAS-QHVIITQCRKRSUKQPLQGEIKFERKVIANDTSTQWKNMNYLAQGTLTQLEY 1364
QY 1377 NEKEKGATITQSPGLDCLTSHSISIPQANRSPLEIAKVSPFPIRPIYITRVLFQDNSSHLP 1436
Db 1365 NEKEKGATITQSPGLDCLTSHSISIPQANRSPLEIAKVSPFPIRPIYITRVLFQDNSSHLP 1424
QY 1437 AA----SVRKQDSGVQBSHFLQAKKNNLSLILTLTQDQREVSGSLGTSATNSVTYK 1492
Db 1425 ASACNYTPRRTSGYQEGSHFLQEAQRNNLSLAFVTLGITGCGKFSGLKSATNQPMYK 1484
QY 1493 KVENTVLPKPDLPKTSVGVKVELLPKVHYIQKOLFPPTETNGSPGHLDLVEGSLAQTEGAI 1552
Db 1485 KIENLVLLQPLGSETSDVKVELLSQVEVQBSFPFKTNSDNPFGHLDLNGKFLPKTQGFV 1544
QY 1553 KWNENRPFKVPFLVATESSAKTSPKLDPLAWDNHYGTQIPKEBWSQKBSPEKTAFK 1612
Db 1545 KWKNTSPGVFFLKWATESSEKIPSKLLGLVLDNDHYDTQIPSEWKSQKSTNTAFK 1604
QY 1613 KXDTILSLNACSNHAIAINBGNQKPLEVTWAKQGTERTLCSONPPVLKEHOREIITST 1672
Db 1605 EXDTTLPLGPCNNDSIAINEGQDKPOEAWAKQGBPRLCSONPPVSKRHQREIIVT 1664
QY 1673 TLQSDOEEDIDYDTISVEMKKEDFDIYDEENQSPESFOKTRHYFIAAVERLWDYGSS 1732
Db 1665 TLQSDOEEDIDYDTISVEMKKEDFDIYDEENQSPESFOKTRHYFIAAVERLWDYGMSR 1724
QY 1733 SPHVLRNRAQSGSVQPKVQFQETDGSFTQPLIRGELNEHGLGLGPVIRAEVDENIMV 1792
Db 1725 SPHVLRNRAQSGSVQPKVQFQETDGSFTQPLIRGELNEHGLGLGPVIRAEVDENIMV 1784
QY 1793 TPENQASRYSFVSSLSIYEEBQROGABPKFVKNETKTYFWKVOHHMA*TKDEBDC 1852
Db 1785 TPENQASRYSFVSSLSIYEEBQROGABPKFVKNETKTYFWKVOHHMA*TKDEBDC 1844
QY 1853 AWAYFSDVLEKDVHSLGILPILVCHTNTLNPAHGRQVTVQFALPFTTFIDETKSYFTE 1912
Db 1845 AWAYFSDVLEKDVHSLGILPILVCHTNTLNPAHGRQVTVQFALPFTTFIDETKSYFTE 1904
QY 1913 NMERNCRACNTQMDPPTKENVRFAINGYIMDTLPGLVMAQDQIRWYLLSMGSENI 1972
Db 1905 NLERNCRACNTQMDPPTKENVRFAINGYIMDTLPGLVMAQDQIRWYLLSMGSENI 1964
QY 1973 HSHFSGHVPTRKKEEYKMAVNLPGVPFVEMLPKSKAGIWRVECLIGEHLHAGMSTL 2032
Db 1965 HSHFSGHVPTRKKEEYKMAVNLPGVPFVEMLPKSKAGIWRVECLIGEHLHAGMSTL 2024
QY 2033 FLVYSNKCQTPLGWASGHIRDFQITASGOYQWAPKLAELHYSGSINAWSTKBPFSWIKV 2092
Db 2025 FLVYSNKCQTPLGWASGHIRDFQITASGOYQWAPKLAELHYSGSINAWSTKBPFSWIKV 2084
QY 2093 DLLAPMIHGIKTQCAROKESSLIYSOFILMYSLDGKKWCTYRGNSGTGLMWFFGNVDS 2152
Db 2085 DLLAPMIHGIKTQCAROKESSLIYSOFILMYSLDGKKWCTYRGNSGTGLMWFFGNVDS 2144
QY 2153 GIKHNFENPILAIYRILHPTHYSIRSLRWELMGCDLNSCMPLGMSKAIISDAQITAS 2212
Db 2145 GIKHNFENPILAIYRILHPTHYSIRSLRWELMGCDLNSCMPLGMSKAIISDAQITAS 2204
QY 2213 SYFTNMFATWSPKARLHLOGRNAMQPQVNNPKEWLOVDQFKTMKVGTGTTQGVKSLIT 2272
Db 2205 SYLSSMLATWSPQARLHLOGRNAMQPQVNNPKEWLOVDQFKTMKVGTGTTQGVKSLIT 2264
QY 2273 SMYVKEFLISSQDCHQWTLFPQNGKVKVFGQOGSETPVWNSLDPLLTRYLRHPOSW 2332
Db 2265 SMYVKEFLISSQDCHQWTLFPQNGKVKVFGQOGSETPVWNSLDPLLTRYLRHPOSW 2324
QY 2333 VHOIALRWEVLGCEAQ 2348
Db 2325 AHHIALRWEVLGCEAQ 2340
```

```
RESULT 3
Q7TN96 PRELIMINARY; PRI; 2258 AA.
AC Q7TN96,
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Factor VIII.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
CX NCBI_TaxID=10116;
RN [1]
SEQUENCE FROM N.A.
RC STRAIN=Wistar; TISSUE=Liver;
RA Katzka M., Geisen C., Seifried E., Oldenburg J.;
RT "Sequence of the rat factor VIII cDNA.";
RL Submitted (AUG-2003) to the EMBL/GenBank/DBSJ databases.
DR EMBL; AY362193; AAQ31580.1; -.
SQ SEQUENCE 2258 AA; 251299 MW; 13AF91C789059B1D CRC64;

Query Match 50.6%; Score 6276.5; DB 11; Length 2256;
Best Local Similarity 53.2%; Pred. No. 0;
Matches 1271; Conservative 317; Mismatches 630; Indels 171; Gaps 29;

QY 1 MQEILSTCFFLCLLRFCFSATRRYYLGAVELSWDMQSDL-GEIPVDAFPFPRVPKSPFF 59
Db 1 MEAARGLCFFLSLCAIASCATRKYYLGAVELPMDYVSGASGACRSDSRFPFGTAP- 57
QY 60 NTSVYVYKTLFWETDHLFNIAKRPWMLGILPTIQAEVYDVTVTILKNASHVPSLHA 119
Db 58 SARVHTTIVFVYMDRPFHTAKRPLMWGLLPTITWTEVHDVTVTILKNASHVPSLHA 117
QY 120 VGVSVMKASGAEYDDQTSQREKEDKVPFGSHTYVMQVLKENGPMASDPLCLTYSYL 179
Db 118 VGMSFWKASGAEYDDHSGSPAEDKDDKVLPGESHTYVMQVLGSGPMASDPLCLTAYLS 177
QY 180 HVDLVKDLNSGLIGALLVCRGSLAKRKTQTLHKILLPAVFDEGSKSHSETKNSLMQDR 239
Db 178 HVDLVKDLNSGLIGALLVCRGSLAKRKTQTLHKILLPAVFDEGSKSHSETKNSLMQDR 230
QY 240 DAASARAPKXHTVNGYVNNRSLPGLIGCHRKSVYHVGITGTTPEVHSITLHGHTFLVN 299
Db 231 DEASTEAOFTVHTVNGYVNNRSLPGLIGCHRKSVYHVGITGTTPEVHSITLHGHTFLVN 290
QY 300 HQASLEISPTIFLAQTLMDLQGLLSCHTSSHQHGMAYVQVDCPEAPQOGKED 359
Db 291 HQASLEISPTIFLAQTLMDLQGLLSCHTSSHQHGMAYVQVDCPEAPQOGKED 350
QY 360 EBAEDYDDLTDSEMDVVRFDNNSPFTQTSVAKKHPKTHVYIAAREEDWDYAPLV 419
Db 351 EBTEDYDD--LDSEMDVFTWDPDAAP-FVQVRSAAKRPKTHVYIAAREEDWDYAPLV 407
QY 420 APDRSKSYQLNNGPQIRGKVKVRPMAYTDETPKTRREALQHESGILGPLLYGEVGT 479
Db 408 TLDEGSLRNLRYLGRGPRIGSKYKVRVYVYDGTFTREATPRAGLILGPLLYGEVGT 467
QY 480 LLTIIFKNQASRYNIYPHGITDVRPLYSRRLPKGVKHLKDFPLPGEIFKYKWTVTVEG 539
Db 468 LLTIIFKNQASRYNIYPHGIRDVGAHVAGRLPRGVKHXDLPIRPGETPKYKWTLTAE 527
QY 540 PKSDPRLCTRYSSPVNMRDLASGLIGFLICVYESVDQGNQTMDSKXNVILFSPVD 599
Db 528 PARSDARCVTRYASAVDPRDLASGLIGFLICVCKESVDQGNQTMDSKXNVILFSPVD 587
QY 600 ERESWYLTENIORFLPNPAGVQLEDPEFOASNMHSHNGVYFDSLSQSLVCLHEVAYWYL 659
Db 588 ENKSWIISNMRRFLDDEBAHVQLQDEFEASNLHSHNGVYFDSLSQSLVCLHEVAYWYL 647
QY 660 SIGAQTDLSVFFSGYTFKEKAVYEDTLTFPSGETVFMNMPGLWILGCHNSDFNR 719
Db 648 SVGAQTDLSVFFSGHTFRRHAYADTLTVFPHSGVTVFWSMNDNPGVWVILGCHNDFNR 707
```

QY 720 GWTALLKVSCKNTGDDYEDSYEDISAYLLSKNNALPRSFQNSRHPSTROKQNFATT 779
DB 708 GWTALLKVSCKDGVSDYGETYEGVLA-PLANDNAVDPSPFQNSNHLHRKKSNA--- 763
QY 780 IPENDIEKTDWFAPHTPMPKIQNVSSDILMLRSGSPTEHGLSLDLQAKAYETSDDP 839
DB 764 -----TEM-----RPLSHPTQHGSPWSDSQGDVDAVRAE 795
QY 840 SPGADSNNSLSMTHFRPOLHSGDMVFTBESGLQLRLNEKLGTTAATLKKLDKVS 899
DB 796 SPAAHNEGPAEAAQLEPHPEKQKALPFS---WLRKRSLVTTMEAKRKKLJLQVQ- 851
QY 900 TSNNLSTIPSDNLAACTDNTSSILGPPSMPV-----HYDSQLDTTLFGKSSSLTE 950
DB 852 -----VSGLED-----DRTAVVA PDDPMAACKAGSGGPPDESSPAALGKKMYPRIR 899
QY 951 SGGPLSSENNDSKLLBESGLMNSQESWGNVSTESGRLPKGRKRAHGPAALLTKDNALF 1010
DB 900 PHGPLSLIEGNRDSNSDSTLMYRLGSPPGDATSWTENRRGLGKRSRHRVAFARGNTLL 959
QY 1011 KVSISLLKTKNTSNNSATNKTTHIDGSLLIENSFVWQILIESDTBPKKVTPLIHDRML 1070
DB 960 -----SDVDEKSHAPGPTSGVNGNTAAVQDITLICTSILEVTPPIHRIIL 1004
QY 1071 MDKNATLRNLNHSNKTTSKKNMEMVQKKEGPIPPDAQNPDMMFFKMLFLPESARMIQR 1130
DB 1005 SDKSATVLRPHRTITSTERKDIRHEKGDVLPQDADTSAPFSEAPFSESTDLKE 1064
QY 1131 THGNKNSNGSGGPPKQIVSLGPKSVEGQNFLEKKNVVGKGFPTKDVGLKEMVFPSS 1190
DB 1065 ANGNSAKPEQEPSEFQGLVIMYIKMENOSFSSEKKNMAGCGGFTKNTGLETVFPRK 1124
QY 1191 RNLPLTLNLDNLHENTNOEKKIOEEIEKKTLLQENVLVPHQHTVCTKFMKNLFLLS 1250
DB 1225 TSVLETTVAKQESGRHNCB-NIPQAVTEKRAPEKCALQVHIATGSKNPPDMFVLG 1193
QY 1251 TRQNVESYDGAAPVLDQFSLNDSNTRTKHTAHSKKEEBS--NLEGNGNTOKIVE 1308
DB 1184 TGQNI-NLHEETVYVPHKNVPTWNTPTDTRQIPMVHFFPKKEBETNSGGLVKNKTREIVR 1242
QY 1309 KYACTTRISNPTSQNFVQSKKALQFRLPLETELEKRIIVDDTSTQWKNMKHLTP 1368
DB 1243 NY-----PSQKNSVARREQASGRK-----ASARLWPDVNRSIQ 1277
QY 1369 STLQCIDYNEKKGATQSPSLDCGLTRSHSIPQANRSLPIAKVSSPESIPVLTVLVLF 1428
DB 1278 SLLKQIDHRKERRKFIIESRRADS-SGTQSTQNHSPSHVWKSAPP---PRDIRIBS 1333
QY 1429 QDNSHLPAASY-----RKDSGVQESSHFLOQAKNNLSLAILTLEMTGDQREVSLGTS 1484
DB 1334 RD-SQVWVSSVAYDFETGSSRIRESSPFLNETDNPDLAVPPRFRVRRGRPASPEKV 1392
QY 1485 ATNSVTYKKNVTVLPKDLPLKTSKVELLPKVHIYQKDLPTETNSNGSPGHDLVBSGL 1544
DB 1393 NTHSAPCDKPNVSVSEFVSPCEAVTALPHVSTQBEELPAGSSCBRRGHVDLFLBIS 1452
QY 1545 LQCTEGATKMEANRPGKVPFLRVATESSAKTPSKLLDPLAWDNHYGTOIPKEBWKQEX 1604
DB 1453 LQRTGPGVERGRGRSGG-----DTGHTENPGKTPSPL-----PQMPKQDGSAG 1499
QY 1605 SP-----EKTAFKKKOTILSINACESNHAIAINEGQNKPEIEVTWAKQKTERLCSON 1658
DB 1500 YPKICRWEDGTVLPFRPHNLSLGA-----KKKNLPRERATRVQGEAPSLVSPE 1549
QY 1659 PVLKXHQREITR-----TTLOSQDEEI--DYDDTISVEMKKEDFDIYDE 1701
DB 1550 PLVLFRRRPREASTLLPGCEMRDDGGVTAEYDDDDDVMAEYDDDAVTVTDPECDYIYGE 1609
QY 1702 DENQSPERSFOKTRHYFTAABELLDYGNSSSPHVLNRASQSGSVQPKVVFQBFDTGS 1761
DB 1610 EAGQPGRFGQKTRHYFTIAABELLDYGNVSPGALGDAWSGDAARFKVVFRTDTGS 1669
QY 1762 FTQPLYRGELNHLGLLOFYIRAEVEDNMVTRNQASRPYSFYSLISYEEDSQSGAPP 1821

DB 1670 FTQVRHGELDAELGLGPIYRAEVEDNMVTRNQASRPYSFYSLISVSPFAD-GGAAP 1728
QY 1822 RKNFVKNETKITVFKVQVHMAPTKDFDCKAWAYFSDVDLEKDVHSGGLIGPLLVCHTNT 1881
DB 1729 RSNFVKNETKITVFWVRPHMAPTKDFDCKAWAYFSDVDLEKDVHSGGLIGPLLVCHAST 1768
QY 1882 LNPAGRQVTVQBFALFFTIPTDKTSWYFETENMERNCRAPCNMQMEDPTPKENYRFAIN 1941
DB 1789 LHSSEARPLAVQBFALLFAVDFETKSWYCAENLDRKCRPRGTPAGDPAMREBYRPHAVN 1848
QY 1942 GYIMDTPLGLVMAQDQIRIWIYLSYSGNSNIHISPHSGHVFTVRKHHYKALINLYPGV 2001
DB 1849 GYVADALLGLVMAEGRRTRWHLLSMGDPGHAQSVHFSASVTVRVDGGEHRTAVCNLYPGV 1908
QY 2002 FETVEMLPKAGVWRVVECLLGEHLHAGMSTLFLVYSNKCTOTPLGMASGHIRDFQITASGQ 2061
DB 1909 FTTVEMLPSPAGVWRVSVLGEHLRAGMSALFLVYSTROVPLMGASGYIRDSQITASGY 1968
QY 2062 YGOWAPKLARLHYSGSINAWSTKEPFSWIKVDLLAPMIHGIKTQGARQKPFSSLYISQFI 2121
DB 1969 HGLWTPQLARLHNAGSVNWSAKPEPAWIKVDLLAPMLHGIETQGARHRLSSLYVSQFI 2028
QY 2122 IMYSLDGKQWLYSGKSTGLSVFFGNVDASTVRHNRFPDPIVARVIRVHPTASRTAL 2088
DB 2182 RMELMGCDLNSCMPZGMEKSAISDAQITASSYFTNMFAFMATMSPSKARLHQGRSNARWQP 2241
QY 2089 RMELMGCDLNSCAPLNGESGWSIDQVTASSHLSITFSKAPFALARLHLAGGNARWQP 2148
DB 2242 VKNPKWLVQDFOKTKMVTGVTGQVKSLTSMYKVEFLISSQDGHQWTLPTQNGKVKV 2301
DB 2149 VNDPTQWLVQDLQRTVKTGVTGQVARSULTAMFVKFLVSTQDGEHRTVHLQDGKVKV 2208
QY 2302 FQGNQDSFTVWNSLDPPLTLRYLRHPQSVWVQIALRMVILGCEAQL 2350
DB 2209 FQGNRDASTPMVNSLSPRFRTRYLRHPQSVWVQIALRMVILGCEAQL 2257
RESULT 4
Q804W6 PRELIMINARY; PRT; 1639 AA.
AC Q804W6; (TREMBlrel. 24, Created)
DT 01-JUN-2003 (TREMBlrel. 24, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE Coagulation factor VIII precursor.
GN F8.
OS Fugu rubripes (Japanese pufferfish) (Takifugu rubripes).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorphia; Acanthopterygii; Percomorpha; Tetraodontiformes;
OC Tetraodontidae; Tetraodontidae; Takifugu.
OX NCBI_TaxID=31033;
RN [1]
RP SEQUENCE FROM N.A.
RA Davidson C.J., Hirt R.P., Ial K., Snell P., Elgar G.,
RA Tuddenham E.G.D., McVey J.H.;
RT "Comparative sequence analysis and molecular evolution of blood
coagulation genes from Gallus gallus and Fugu rubripes";
RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF465279; AA033374.1; -
DR GO; GO:0005507; F: copper ion binding; IEA.
DR GO; GO:0007155; P: cell adhesion; IEA.
DR InterPro; IPR001117; Cu-oxidase.
DR InterPro; IPR008972; Cupredoxin.
DR InterPro; IPR000421; FA58 C.
DR Pfam; PF00394; Cu-oxidase; 2.
DR Pfam; PF00754; F5 F8 type_C; 2.
DR SMART; SM00231; FA58C; 2.
DR PROSITE; PS01285; FA58C_1; 2.


```

1296 LVCTVGEISLQWRAKLLVNFCCSPRLGKSGRICDSQIKASDYIGNLPLHARLDQS 1355
QY 2076 GSNIAWSTKPFPGWIKVLDLAPMIHGIKTQGAQKFSYIYIQFIIMYSLDGKQWQTYR 2135
Db 1356 GYINAWGYNEKSWIQVDLQAPTLHRVQIQGVRSNLRNNYITAFYYSLDQETWSTYR 1415
QY 2136 G-----NSTGTLNVFNGVDSSIKENINFPPIIARYIRLHPHYSTIRSTLAMELGCD 2189
Db 1416 GSGSSRSSSSSTAKVFNGLDNRVKNPFPVPPVARYIRIHPHYSTIRSTLAMELGCD 1475
QY 2190 LNSCSMPLGKESKAISDAQITASSYFTNMFATWSPSKARLHLOGRSNAWRPQVNNPKEWL 2249
Db 1476 LNSCSLPGLQDRIPRESFVASSYSWSLRSWTFPSLARLHQEGSANAWRPKNNPHEWL 1535
QY 2250 QVDFQKTMKVTGTQGVKSLTSMYKVEFLISSQDGHQWTLFFQNG--KVKVFQGNQD 2307
Db 1536 QVDLGKVKRITGVVGTQARSLLTRNMVTEFVTTISRDQAMSSVLEGSSOREKIFQGNND 1595
QY 2308 SFTPVNSLDPLTLRILRIHPQSVWVHQAIAEMEVLGCEAQ 2348
Db 1596 SDEEALTIFDAPLGRYIRIHPGLWINDIALRLEVILGCDTQ 1636

RESULT 5
Q90X47 PRELIMINARY; PRT: 2119 AA.
AC Q90X47;
DT 01-DEC-2001 (TREMELrel. 19, Created)
DT 01-DEC-2001 (TREMELrel. 19, Last sequence update)
DT 01-OCT-2003 (TREMELrel. 25, Last annotation update)
DE SC:b22015.3 (Novel protein similar to vertebrate coagulation factor V
DE and VIII).
GN SC:B22015.3.
OS Brachydanio rerio (Zebrafish) (Danio rerio).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC Cyprinidae; Danio.
OX NCBI_taxid=7955;
RN SEQUENCE FROM N.A.
RA Lloyd D.;
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: CONTAINS 2 F5/8 TYPE C DOMAINS.
DR EMBL; AL590146; CAC94896.1; - binding; IEA.
DR GO; GO:0005507; F: copper ion binding; IEA.
DR GO; GO:0007155; P: cell adhesion; IEA.
DR InterPro; IPR001117; Cu-oxidase.
DR InterPro; IPR008972; Cupredoxin.
DR InterPro; IPR000421; FA58 C.
DR InterPro; IPR008979; Gal_Bind_like.
DR Pfam; PF00394; Cu-oxidase; 2..
DR Pfam; PF00754; F5_F8_type_C; 2.
DR SMART; SM00231; FA58C; 2.
DR PROSITE; PS01285; FA58C.1; 2.
DR PROSITE; PS01286; FA58C.2; 1.
DR PROSITE; PS00022; FA58C.3; 2.
DR PROSITE; PS00079; MULTICOPPER_OXIDASE; 3.
SQ SEQUENCE 2119 AA; 240643 MW; DCE806FFA8761E6 CRC64;

Query Match
Best Local Similarity 30.3%; Pred. No. 2.2e-159;
Matches 752; Conservative 379; Mismatches 827; Indels 525; Gaps 68;
QY 11 LCLRF-CFSAT---RRYVLGAVELGWDYQSDLGELFVDARPPRPVPSFFNTSVVYK 66
Db 13 LALLAEFCATAVRHHVYIAAVNINWDTSGQQ-----RTQGSYK 53
QY 67 KILFVFTDHLFNIARPPRMGLLGTQCAEYVDTVITLKNWASHPYSLHVGYSYWK 126
Db 54 KVIYREYNEG-FKQPKAHLPSGLGLTUKRGQGGDTIIIVFRNADHPCLSPHFGAYGK 112
QY 127 ASGCAEYDQGT-SQREXEDDKVFPGGSHYVWQVLKENGPMASDPLCLITYSLSHVDLVXD 186

```

```

113 QSSGSLVFNSTLSLEKNDVIOQGESEHTYQWDVTSVDTPTAADPPCITYTLSHFDIVRD 172
QY 187 LNSGLTGALLVCHREGSLAKKKTQTLH---KFILLFAVEDSGKSWHSETKNSLQMDRDAAS 243
Db 173 YNTGLIGEMLIICKGTLDSSGNO-IHFHQESVLLFGVFDENKSWYS-----TGDS 221
QY 244 ARAPKMHVTVNGVNRSLPLIGICHRKSVYVHVIGMTTPEVHSIFLEGHTFLVRNHRQA 303
Db 222 PQPLNVKYTINGYNGSVPLDICAHSKVSWLLGMSSEPELFSVHFNGQVLLHDGKTS 281
QY 304 SLBISPTFTTAQTLMLDGLQFLLSCHISHQHDGMEAYKVDSCPEPOLPMKNNEAE 363
Db 282 AVGLISGTSATSASMTGVHPGRWLVSSSHISKHLAAGLHGYLNIRKC----- 326
QY 364 DYDDDLTDSMDVVRPDDNSPSPFIQIORSVAKKHPKTVWHVIAAEEDWDYAPILVLPDD 423
Db 327 -----DEYAPK-RLTIQKESQWYMAAEVINDYAPNKPENMD 369
QY 424 RSVKSOYLNNGPORIGRKYKVRPMAYTDTFKTR---EAIQHSGLGLLYGEVDTL 480
Db 370 GDFESKYLKQGPQRIKGYKKAFTQYKDGWFKERABDKQRKRELGLPGVIRAIQIRDI 429
QY 481 LIIPKQASRPYNTYPHGITDVRPLYSRRLPKGVKHLKDPILPCEIFKYKWTVTVDGP 540
Db 430 KIVFNKASRPYSITYPHEGLTIDKAAEGASYPQGN--QYTSVQPGETYTYTWSVTEEDVP 487
QY 541 TKSDPRGLTRYSSFVNMERDLASGLIPGLICYKESVDQKNGQIMSDKRNVLIFSVEDE 600
Db 488 TDSDFCLTRMYHSADVAPRIASGLVGPLLICHSQSNKKNVQKADKEQHAFMTVDE 547
QY 601 NRSWYLTENIORFLNPAGVQLEDPFQASNMHSINGYVDSIQ--LSVCLHVEYAVYWL 659
Db 548 NKSVMQDENINTYCSDPKVKKDDPPEFYKSNVMTINGYVYESGQELGFCHEIVTWEVS 607
QY 660 SIGAQTFPLSVFPGYIFKHMVVEDTLTLPFPGSETVFMSEWEPGLWILCHNSDPRR 719
Db 608 SVGEQDYQTATFYGHTFELKNREEDILSLPMTGETITMNVNIGIWLASLNSHDSK 667
QY 720 GMTALLKVSCKNTGDY--YEDSVEDI-SAYLLSKNAIEPRFSQNSRHPSTROKFNAT 778
Db 668 GMVKFKDLECFR---DYVIDYED-----GKFTAKWPTI----- 701
QY 779 TIPENDIEKTPWFARHTPMPKIKNVSSDILMLLROSPTPHGLSLDLQAKYETFSDD 838
Db 702 ---NEIKKEEPVAFAR----- 713
QY 839 PSPGAI DNSNSISEMTHFRPOLHSGDMVFTPESGQLRLNEKLTGTTAATLKLKDFKVS 898
Db 714 --PQVVEYSDLFZET-----LNRFTNVVKDEVEIIDTLFL 748
QY 899 STSNLI STIPSDNLAAGTNDNTSSILGPPSPMPVHYDSQLDTTLFGKKSPLTESGGPLSL 958
Db 743 DODDGLLPIVEKSLGSSNEN-----LHNATLQSFIEHGO----- 783
QY 959 BENDSKILLEGSLMNSQSSWCKNVSTESGRLFKGKAHGAHPALLTKNALFKVSIILK 1018
Db 784 -----LLMEGDLKGESSKNVLNDSTK-----ALLE 811
QY 1019 TNKTSNNS---ATNRKTH---IDGPSL--LIENSPSVQNILESDFEKKVTPLEIHDRML 1070
Db 812 TTTTFDSNRVVALNNETDSIILDFFIVERKVRSAFSKPMNEPESVTMFKITEHI----- 866
QY 1071 MDKNATARLNMNSKNTTSSKNMEMVQOKKGPIPP-DAQNPDMSF-----FKMLFLPES 1124
Db 867 ---NSSLERINAI:SPITEFNINTWETHTDPSITPFDGSTGEMFTLEDDTALLNSSES 923
QY 1125 ARWIOETHGKNSLNSQGPSPKQLVSLGPEKSVBQGNF-----LSEKNVYVVGKE 1175
Db 924 EPLQNSONSENRIAFQELNAKDGTDVDSNNNSVKNQIKFKNVPSGDTLSSNSKIQV---- 979
QY 1176 FTKDVLGKEMVFPSSRNFLTNLDNLHNNTNNOEKKIOEBIEKKEKTELLQENVLPQIHT 1235

```

```
Db 980 -----BEDFVLLDSSYFSBMSST-----TMEYDVQSKDTVKGESKETAQSQE 1020
Qy 1236 VTGKNFMKNFLPLSTRONVEGSDYGAAPVLQD-----FRSLN-----DGTN 1278
Db 1021 LSSYTKKTSYGBILLESPLDITSAFNLNDSVLRNNSLSNSSNETLFWSNAPFSDSTN 1080
Qy 1279 RTKKHT-----AHFKKG-EEENLEGLNQTKQIYEKACTRIGFNTSQQNFVQSKR 1332
Db 1081 ATSSDSSATFADFNTTFSNATPDSFNRIQMSDSSNATLSDSSNAT-----LSDSSNA 1136
Qy 1333 ALKQFRLPLEETELKRIIVDTSTQWKNMKHLTPS-----TLQIDYN-- 1377
Db 1137 TLSD-----SSNATLSDSSNATLSDSSNATPDSNKTFSNATLFGVSYSSN 1186
Qy 1378 -----EKEGATQSPSLDCLTRSHSIPQANRSPPIAKVSSFPSPRIPIVLRVLFQ 1429
Db 1187 TTLSFSLSESTEYILSSANDTI-KSHSEVVSNTSOL-----SSSESTENISL---LYG 1237
Qy 1430 DNSSHLPAASYRKSGVQESSHFTLOGAKXNLSLAILFLEMTGQREYVSGLSGTATNSV 1489
Db 1238 SLN-----ASSMKNDSESEBEVVIYNKNH-SEAILTSHL--DOKE-EHWGYESKHEL 1288
Qy 1490 TYKKVENTV-----LKPDLPTSGKV--ELLPKVHI-----YQKDLFTEISN 1531
Db 1289 VHELFDHVNKYVKDKSAANSNKPXIEKXKVIQKVKPKGYGMKTKSKDYKQPPRSS 1348
Qy 1532 GSPGHLDLVEGSLLOQTEGAIKWNANRPKVPFLRVATESS-AKTPSKLLDPLAWDHY 1590
Db 1349 FSP-----RGFGSVLTTPRSRP-----VSSSEDELTEKPIVIGVPRDFNDY 1390
Qy 1591 GTQIPKEEKWSEKSEKTAPEKXKOTILSINACESNHAIAINEGKNKEIETVAKQCR 1650
Db 1391 ELYIPKQDOEAD-----FDGLL-----DHPE----- 1411
Qy 1651 TELSCQNPPVLRKHQREHTRTLQSDQBEIDYDDTISVEMKKEBFDIYD-EDENOSPRS 1709
Db 1412 -----EYEXVEKQPYSS-----KTADVQALDATSQHLLKWA 1442
Qy 1710 FOKKTRHYTIAVERLDYGMSSPHVLRNRAQSGVPOFKKVVQFEFTDGGFTQPLYRG 1769
Db 1443 GDANTRTYISVEEENDYA-GVQRRLDKTAQNERPTVFRKVVFRYLDSTFIRDING 1501
Qy 1770 ELNEHLGLGPYTRAIVENIMVTFNQASRPYSFYSSLISY-----BEDQOGAEP 1821
Db 1502 EMDEHLGILGLPLKAEVDQTMVVFERNASRPYSLHANGVKYKQMEGLSYDDESPYWK 1561
Qy 1822 RKNFKVNETKTVPKVKQHUMATKDBPCAKAWAFSDVDLEKDVHSGGLGLLVCHTNT 1881
Db 1562 QDDAVPNTGTTTMTNTINKSPQNNECDRTWTYSAVNFERDINSGLLGLLVCKGT 1621
Qy 1882 LN--PAHGRQVTVQEPALFTTIEDETKSWYFTENMERNCRAPNIQMEDPTFKENYRFA 1939
Db 1622 LDKKPEDRR-----EVLLFMFTDENKSMLEYENRQRIERKNRVVY-DPNFQDLKFDA 1675
Qy 1940 INGYINDTLPCLWAQDQRIRWLLSNGSNENTHSIFSGHVTVEKKEBKALNLYP 1999
Db 1676 ING-ITYSLKGLRMYTNQLAKWHLINSPKDLHSHVHEHQTQINKELKDHQGVPLLP 1734
Qy 2000 GVPEVTEMLPSKAGINRVECLIGEHLHAGMSTLP-----LVISNKKQ 2041
Db 1735 GGFATLEMLPSKGLQWLESEVGLSQORGQMTFLFDESYYKRVCOVLSYLIYFAVCD 1794
Qy 2042 TPLGMSAGHIRDQITASGOYGWAPKAPLARLHYSGSINAWST-KEPFSWKVLLAPMII 2100
Db 1795 HPLGLISGTQDEQITASDRGQWYPHPLASLHNTGKYNASTSSEPQYQLQVDFQRPVI 1854
Qy 2101 HGIKTOGAROKFSLYSIQFIIMVSLDGKQWQYRGNSTGT-----LMVFF----- 2146
Db 1855 SKVATOGAKOFLTHFNFLNTISYSTDKKWIYKGDSDAVRKANIHKTNTVFNARQTFE 1914
Qy 2147 GNVDSGGIKHINFPPIIARIHLRPHYSIRSTRFLRMELMCDLNSCMPLGVSKAISD 2206
Db 1915 GNAEAYETKENIFPPPLIGRIVLRPLHSHYNTFVIRLEYGCELDGCGSVFLGKEKGLIDD 1974
```

```
Qy 2207 AQITASSYFTINMEA-TWSPKARLELQGRSNAPRPOVNNPKEWLQVDFQKTMKVTGVTQ 2265
Db 1973 SKITASSVASNWYSGQHPWYARLNKQCTANAWQAKNDIQPWIQVELKEVKKITGIVTQ 2034
Qy 2266 GVKLLSTMYKEFLISSQDCHOWTLFFQNG--KVKVFGQNDSTFPVNSLDPPLLTR 2323
Db 2035 GAKSGMNGEYFVRSYILEYSESGREWMKYTDDEQYKLFQGNTDNNGQIKNIYIYFIFSR 2094
Qy 2324 YLRTHPOSVWHQIALRMEVLQCE 2346
Db 2095 FIRIIPKQWQSVTRIELLQCD 2117
RESULT 6
O88783
ID O88783 PRELIMINARY; PRT; 2183 AA.
AC O88783;
DT 01-NOV-1998 (TREMELrel. 08, Created)
DT 01-NOV-1998 (TREMELrel. 08, Last sequence update)
DT 01-OCT-2003 (TREMELrel. 25, Last annotation update)
DE Murine coagulation factor V.
GN F5.
CS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98282202; PubMed=9616155;
RA Yang T.H., Cui J., Rehmulla A., Yang A., Mousalli M., Kaufman R.J.,
RA Ginsburg D.;
RT "The structure and function of murine factor V and its inactivation by
RT protein C.";
RL Blood 91:4593-4599 (1998).
CC 1- SIMILARITY: CONTAINS 2 F5/8 TYPE C DOMAINS.
DR EMBL; U52925; AAC99553.1; -.
DR PIR; T42764; T42764.
DR RSP; P12259; 1CZT.
DR MGD; MGI:89382; F5.
DR GO; GO:0005507; P:copper ion binding; IEA.
DR GO; GO:0007155; P:cell adhesion; IEA.
DR InterPro; IPR001117; Cu-oxidase.
DR InterPro; IPR008972; Cupredoxin.
DR InterPro; IPR000421; FAS8C.
DR InterPro; IPR008979; Gal_Bird_like.
DR Pfam; PF00394; Cu-oxidase; 3.
DR Pfam; PF00754; F5_F8_type_C; 2.
DR SMART; SM00231; FAS8C_2; 2.
DR PROSITE; PS01285; FAS8C_1; 2.
DR PROSITE; PS01286; FAS8C_2; 2.
DR PROSITE; PS00022; FAS8C_3; 2.
DR PROSITE; PS00079; MULTICOPPER_OXIDASE1; 2.
DR SEQUENCE 2183 AA; 247228 MW; BFOA8AA723F60317 CRC64;
```

Query Match 22.3%; Score 2771.5; DB 11; Length 2183;

Best Local Similarity 29.8%; Pred. No. 3e-156;

Matches 749; Conservative 409; Mismatches 851; Indels 503; Gaps 73;

```
Qy 8 CFPLCLL---RFC-----FSATRRYYLGAVELSWDYMQSDGLGELPYDARFPVRPKS 56
Db 7 CFLLVVLGTRWAGWSHQAEAAQLQFYVAAGILWYHPE-----PTD 51
Qy 57 FPFNTSVVYKTLFVEFTDHLFNIAKPPFPMGLGPTIOAEVDTVITLKNMASHPVIS 116
Db 52 PSLSNTPSKVIYREY-EQYFKKEKPRSSNGLLGPTLYAEVGVIVKHFFENKADKPLS 110
Qy 117 LRAVGSYVKASEGABYDDQTSQREKDDKVPFGSGHSTVVOVLKENGPMASDPLCLTYS 176
Db 111 IHFQGIKYSKFSEGAASYADFTFAERKQDAVAFGEYTYEWISVDSGPTDPDPCLTTHI 170
Qy 177 YLSHVDLVLDKSLGALLVCRGSLAKEKTQL--HKFILLFAVDFDEGKSHSSTKNS 234
```



```
QY 2242 VNNPKEVLQVDFQTKMKTGVTGTTQGVKSLLTSMYVKEFLISSQDGHOWTLFFQNGKV-- 2299
Db |||:||||| |||:||||| |||:||||| |||:||||| |||:||||| |||:|||||
2075 ANNKKQWLQVLLKLLKVVTAIVTGCKSLSEMVKSYISQGVAKKPYRKSSWVD 2134
QY |||:||||| |||:||||| |||:||||| |||:||||| |||:||||| |||:|||||
2300 KVFQGNQDSFTPVNSLDPPLRLRYLRHPSWVHQIALRVVLGCRADQIY 2351
Db |||:||||| |||:||||| |||:||||| |||:||||| |||:||||| |||:|||||
2135 KIFGNSNTKGMKNFFNPPIISRFIRIIPKTNQSTALRLRFGC---DIY 2183

RESULT 7
Q804W5 PRELIMINARY; PRT; 1802 AA.
ID Q804W5;
AC Q804W5;
DT C1-JUN-2003 (TREMELrel. 24, Created)
DT C1-JUN-2003 (TREMELrel. 24, Last sequence update)
DT C1-OCT-2003 (TREMELrel. 25, Last annotation update)
DE Coagulation factor V precursor (Fragment).
GN F5.
OS Fugu rubripes (Japanese pufferfish) (Takifugu rubripes).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
OC Tetraodontidae; Tetraodontidae; Takifugu.
OX NCBI TaxID=31033;
RN [1]
RP SEQUENCE FROM N.A.
RA Davidson C.J., Hirt R.P., Lall K., Snell P., Elgar G.,
RA Tuddenham E.G.D., McVey C.H.;
RT "Comparative sequence analysis and molecular evolution of blood
RL coagulation genes from Gallus gallus and Fugu rubripes.";
RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF465280; AA033375.1; -.
DR GO; GO:0005507; F:copper ion binding; IEA.
DR GO; GO:0007155; P:cell adhesion; IEA.
DR InterPro; IPR001117; Cu-oxidase.
DR InterPro; IPR008972; Cupredoxin.
DR InterPro; IPR000421; FAS3 C.
DR InterPro; IPR008979; Gal Bind like.
DR Pfam; PF00394; Cu-oxidase; 1.
DR Pfam; PF00754; F5 F8 type-C; 2.
DR SMART; SM00231; FAS3C; 2.
DR PROSITE; PS01285; FAS3C_1; 2.
DR PROSITE; PS01286; FAS3C_2; 1.
DR PROSITE; PS00022; FAS3C_3; 2.
DR PROSITE; PS00079; MULTICOPPER_OXIDASE1; 2.
FT NON_TER
SQ SEQUENCE 1802 AA; 205160 MW; B2EC3D1D9AA8367D CRC64;

Query Match.
Best Local Similarity 30.3%; Pred. No. 3.2e-151;
Matches 709; Conservative 323; Mismatches 719; Indels 586; Gaps 67;

QY 62 SVVYKTLFVEVTDHLFNIAKRPMPWGLGPTIOAEVYDVVITLKNMASHPVSLHVG 121
Db |||:||||| |||:||||| |||:||||| |||:||||| |||:||||| |||:|||||
2 SPYRKVVPREY-DKDFQPKRPPWGLGPTLRAEBGEIIVTFRNLAATPKYSIHPHG 60

QY 122 VSWKASEGAZDDOTSQREKDDKVFPGGSHTYVWQVLKNGPMASDPLCLTYSYLSHV 181
Db |||:||||| |||:||||| |||:||||| |||:||||| |||:||||| |||:|||||
61 VAYKQSEGANYFDNTSQEKEDDVVPNSHVYWEITSDVSPQNDPTCLTYIYISHK 120

QY 182 DLVKNLNSGLIGALIVCRGSLAKETQ--TLHKFILLFAVDFEGKSWHSETYNSLMQDR 239
Db |||:||||| |||:||||| |||:||||| |||:||||| |||:||||| |||:|||||
121 DVVEDYNSGLIGALLVCKSGSLDESQRQIGIVHYVFLFGVFNENE-----SKFKQ 172

QY 240 DAASAAWPMHTVNGYVNSRLPGLICGCKHSYVMEVIGMTTPYVHSIFLEGHFLVFN 299
Db |||:||||| |||:||||| |||:||||| |||:||||| |||:||||| |||:|||||
173 DPASVD--HKYTINGFTGSLPDVSICTYAPVSLHLVGMSSDPVFVSHINGQVTLQNG 230

QY 300 HROASLEISPIITLTAQTILMDLQGLFLSCHSHCHDGMAYVYKVDSCP--EEOPLIMK 357
Db |||:||||| |||:||||| |||:||||| |||:||||| |||:||||| |||:|||||
231 HXNSSVGLISGSSITVSMVAHVHTRMLLSQIMKHITQKGYGVFNNEICDGFAPORTLT 290

QY 358 NNEEAEDYDDDLTDSMDVVVFPDDNSPSFIQIRSVAKKHPKTNVHYIAAEESDWDYAE 417
```

```
Db 291 REQ: |||:||||| |||:||||| |||:||||| |||:||||| |||:|||||
QY 418 VLAPDDBSYKSOVLNNGPQIBGRYKVKVRYMAYTDEFTK--REAIOHESGILGPLLYG 474
Db |||:||||| |||:||||| |||:||||| |||:||||| |||:||||| |||:|||||
317 BPAHDODYKLOVLRQSSRTIRGGYKAVVTLTYNESFTQVHKQKQKDEGLIGPVIRA 376
QY |||:||||| |||:||||| |||:||||| |||:||||| |||:||||| |||:|||||
475 EVGDTLIIIFRNOASRPYNIYPHGITDVRPLYRRRLPKGVKHLKDFPILGEIFKYKWTY 534
Db |||:||||| |||:||||| |||:||||| |||:||||| |||:||||| |||:|||||
377 QIRDVIKIVKKNATREYSIYPHGLTIEKSEGVNYPGSGNH--SHSVQPGETHYTWKV 434
QY |||:||||| |||:||||| |||:||||| |||:||||| |||:||||| |||:|||||
535 TVEDGPTKSPRLCTRYYSFVNMERDLASGLIGPLLI CYKESVDQXGNQIMSKRVIL 594
Db |||:||||| |||:||||| |||:||||| |||:||||| |||:||||| |||:|||||
435 VEEDEPLDDARCLTRYHSAVDTPRDIASGLICKESLNVNRVQLRADKEQHAM 494
QY |||:||||| |||:||||| |||:||||| |||:||||| |||:||||| |||:|||||
595 PSVFDENRWYLTENIQRFLEPNPAGVQLEDPEFOASNIMESINGVYFDSLQ--LSVCLHEV 653
Db |||:||||| |||:||||| |||:||||| |||:||||| |||:||||| |||:|||||
495 FAVFDENYSNYLLDNI--RYRCEHSKVNKADPDFKYSNVMSHINGVYFESSVLFQNGEV 553
QY |||:||||| |||:||||| |||:||||| |||:||||| |||:||||| |||:|||||
654 AYWYILSIGAQTDFLSVFFSGYTFKHKMVEDTLTLFFSGETVFMGMENPGLMILGCHN 713
Db |||:||||| |||:||||| |||:||||| |||:||||| |||:||||| |||:|||||
354 ATWHSVSGAQDYIQTATFYGHTFELNERTEDFLSLYPMTGETTSMNMDNIGVWLLASLN 613
QY |||:||||| |||:||||| |||:||||| |||:||||| |||:||||| |||:|||||
714 SDFRNRGTALLKVSSCDKNTGYEDSYEDI SAILSKNAIEPRFSQNSRPSIRQK 773
Db |||:||||| |||:||||| |||:||||| |||:||||| |||:||||| |||:|||||
614 SHGTTKGMVKFQVECYRD---YQYEYDD-----STNV 644
QY |||:||||| |||:||||| |||:||||| |||:||||| |||:||||| |||:|||||
774 QFNATTIPENDIEKTDPMFAHRTPMPKIQNVSSDMLLRQSPTPHGLSLDLSQEAKE 833
Db |||:||||| |||:||||| |||:||||| |||:||||| |||:||||| |||:|||||
645 EFNV-----N-----NPLSLDTIQE--- 659
QY |||:||||| |||:||||| |||:||||| |||:||||| |||:||||| |||:|||||
834 TFSDDPSPGAIDSNNSLSEMTFRPQLHSGDMVTFPESGLQLNKLGT-TAATELKK 892
Db |||:||||| |||:||||| |||:||||| |||:||||| |||:||||| |||:|||||
660 ---NEKPFILVENKN--GEFDSYTEDLAN-----ELGLRSLKQQNSGSDVEQLDLISF 705
QY |||:||||| |||:||||| |||:||||| |||:||||| |||:||||| |||:|||||
893 LDFKVSSTSN---NLSTIFSDMLAAGTNTSUGPPSPNVHVDSQDITLFGKSSPL 948
Db |||:||||| |||:||||| |||:||||| |||:||||| |||:||||| |||:|||||
706 LDYEVVDVSNETQNFTKT-KNCPFAAEINGLN-----DT----- 739
QY |||:||||| |||:||||| |||:||||| |||:||||| |||:||||| |||:|||||
949 TESSGPLSLSEENDSKLLEGLMNSQSSWGMKNVSTESQRLFKGRA-----HGP 1000
Db |||:||||| |||:||||| |||:||||| |||:||||| |||:||||| |||:|||||
740 -----SISRLVQNLTGADLN-QQINWNSIVMQNNSTLAFNLHVSNEIKLWNHSP 791
QY |||:||||| |||:||||| |||:||||| |||:||||| |||:||||| |||:|||||
1001 ALLTKDNALPKVSGISLLKNTKNTSNNSATNRKTHIDGPSLLIENSPSVQNTILESTEF-K 1059
Db |||:||||| |||:||||| |||:||||| |||:||||| |||:||||| |||:|||||
792 Q--TKNT--FSVIDTDLAAAATVANTISA-----LSVDNTYSK 824
QY |||:||||| |||:||||| |||:||||| |||:||||| |||:||||| |||:|||||
1060 KVTPLIHDSMLYDKNATRLNLNHNKNTSSK-NMWMYQKKEGPIPPDAQNPDMSPFKM 1118
Db |||:||||| |||:||||| |||:||||| |||:||||| |||:||||| |||:|||||
825 BITNLA--GTLQGNLTA-----SDSNVTVSVSRINM-----PATRIP----- 859
QY |||:||||| |||:||||| |||:||||| |||:||||| |||:||||| |||:|||||
1119 LPLPESARWIQTH--GKNSLNSGGQSPKQLVSLGPEKSVGQNFLEKNKVVGKGBFT 1177
Db |||:||||| |||:||||| |||:||||| |||:||||| |||:||||| |||:|||||
860 -----SCLSVSKAFVKGN--NSSDGTNSSLRAHGP-----YLNSSGKV----- 896
QY |||:||||| |||:||||| |||:||||| |||:||||| |||:||||| |||:|||||
1178 KDVGLEKMWFPSSRLFLTNLNHLNHNTHNQEKKIQEEIEKKETLIOENVVLPQIHVT 1237
Db |||:||||| |||:||||| |||:||||| |||:||||| |||:||||| |||:|||||
897 -----NPFISKGNVAALVLRGSSVTVKLPMSKEEL----- 927
QY |||:||||| |||:||||| |||:||||| |||:||||| |||:||||| |||:|||||
1236 GTFKNFKNLPFLLSFTQNVGSGYDGAYAPVLODFRSLNDSTNETKHTAHFSKKGEENLE 1297
Db |||:||||| |||:||||| |||:||||| |||:||||| |||:||||| |||:|||||
928 -----DNTSLKNPSNKTSSSS----- 944
QY |||:||||| |||:||||| |||:||||| |||:||||| |||:||||| |||:|||||
1298 GLGNQTKQIVEKYACTTRISFNTSQCNFVQTSKEALKQFLPLBETELEKRIIVDDTST 1357
Db |||:||||| |||:||||| |||:||||| |||:||||| |||:||||| |||:|||||
945 -----TPSVTNDIWTSSKE-----LSSSESSKEVF----- 971
QY |||:||||| |||:||||| |||:||||| |||:||||| |||:||||| |||:|||||
1358 QWSKNMKHLTPETLQIDYNEKEKGAITQSPLSDCLTRSHSIPQANRSPLPITAKVSSPFS 1417
Db |||:||||| |||:||||| |||:||||| |||:||||| |||:||||| |||:|||||
972 -----IYVKOKKAGLI----- 982
QY |||:||||| |||:||||| |||:||||| |||:||||| |||:||||| |||:|||||
1418 IRPIYILTRVLFQDNSSHLFAASYRKDKSGVBOSSH--FLOGAKKNLSIALITLMTGQ 1475
```

```
Db 983 -----KTSYKTSCHNWTYDGTGTHK-----IVSAEIPDDM 1011
QY 1476 REVSGSCTSATNSVTVKXZENTVLPKDDLP-----KTSCKVELLPKHVIYQKDLPTTET 1529
Db 1012 KKYFEM-SPQTNKKTKVN-----REHFRQKGGMKTKRKYKPKQ-----PSGULPSL 1061
QY 1530 SNGSPGHLDAVEGSLLOGTEGAIKWNEANRP-KGVPLRVATESSAKTPSKLDPLAWN 1588
Db 1062 RGNP-----LMSPGA-----RPGLOPVNN--DETLMVPVIGVPREDFS 1102
QY 1589 HYGQIPIKEEMKSOEKSPKTAFAKKDKTILSLNACESHAIANEQONKPEIEVTVWAKQ 1648
Db 1103 DYELYLPGE-----1112
QY 1649 GRTERLCSQPPVLKXHQREITRITLOSDEIDYDITISVEMKKEPDIYDEENOSPR 1708
Db 1113 -----PDHLVDQPNVK-----ANEYEVNFKOPYRSNEDAKLHLHQTKYLENR 1158
QY 1709 SFQKTRHYFIAAVERLWDYGMSSSPHLNRRAQGS-----VPOFKKVVFQEFDTGSETQP 1765
Db 1159 --DKWRTYFIAAEVQWDY-----AGYQRRRBPQGQVIRHTKIKVVFRLYLDSSERTP 1212
QY 1766 LYRGELNEHGLGPGYIRAEVEDNIMVFRNQASRPYSVSSLSIYBEDQOQA--EPRK 1823
Db 1213 EVRGEVDEHLGILGVIAKAEVGQTIMVVFKNASRP--SLHPNGVSYSK-QTEGLSYEDGS 1271
QY 1824 NF-----VKPNETKTYFWKVQHMAPTKDEFPCKAKAYPSDVLEKDVHSGLIGPLV 1876
Db 1272 NYWYKYNDEVPQGAFTYIMVNVNMPVQTSDESNCRTWAYSYGVNPERDHSGLIGPLV 1331
QY 1877 CHTNTLPAHQGVQV--VOEPALFTTIFDPTKSWFTTERNMERNCEAPCNIQVEDPTFKENY 1935
Db 1332 CQKGTIN-----QELINTREFMLLEFWTPEOSQWFTDRNREIMLRSMR-KWMDPDMENL 1366
QY 1936 RPHAINGYMDTLPGLVMAQOORIRWLLSGNSNETHSIHFSQHVFTVRKK-BEYKAL 1994
Db 1387 KFHSGINGIIVK-LKGLRMYTNQLVSWHLINMGSPKDFNSVHFHGQTFLLHKKKTSYRQAV 1445
QY 1995 YNLPGVETVEMLPKAGIWRVECLIGEHLHAGMSTLFLVYSNKCQTPGLMASGHIRDF 2054
Db 1446 YPLLPGPSFATLEMPKPLGQLETEVGINQKGMQTLFLVLADDYHPLGLSGSVKND 1505
QY 2055 QITASQOQWAPKLARLHYGGSINAWTKPFPMSIKVDILAPMIHQIKTQGARQKPS 2114
Db 1506 QITALTNEGWEPLARLHNQKYNWTDQNSYMIQVDFQRPVVISQVATQGAQFPQA 1565
QY 2115 LYIGOFIMYSLDGKKWQTYRGNSTGLMWFGNVDSGKIKNIPNPPIIARYIRLHPTH 2174
Db 1566 QVSSYIYSNDRRSWSFYAKGDRDDIKVFTGNPNPYDKKNTFPPLIGRIFRHFPLK 1625
QY 2175 YSIRSTLWELMGDLNCSMPGLGSKAIDAOITASSYFTNMPA--TWSPSKALHLQ 2233
Db 1626 WYNKATLMDYGCGLDGCSPVPLGSGHIEDHQITASTASKVSGTWRFLGLNKEG 1685
QY 2234 RSNARPOVNNPKELQVDFORTKMTVGTVTGQVKSLLTSMYVREFLSSQDGHQWTLF 2293
Db 1686 TINAKQAYNDWDQWLQVELQVKKITGIVTQGAFLGAEFVTSFYSHDGNWHPY 1745
QY 2294 FONGKV--KVFQGNODSTFPVNSLDPLLLTRYLRIHFQSWHQIALRNEVLGCEAQ 2348
Db 1746 TDDDGVPKIFMGNTNNDHVNKVMYP2IFGRFIRVPSWITSGTIPMEPLGDCDE 1802
RESULT 8
Q804X3 ID PRELIMINARY; PRT; 1377 AA.
AC Q804X3;
DT 01-JUN-2003 (TREMELrel. 24, Created)
DT 01-JUN-2003 (TREMELrel. 24, Last sequence update)
DT 01-OCT-2003 (TREMELrel. 25, Last annotation update)
DE Coagulation factor VIII precursor (Fragment).
GN F8.
OS Gallus gallus (Chicken).
```

```
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
SEQUENCE FROM N.A.
RA Davidson C.J., Hirt R.P., Lal K., Snell P., Elgar G.,
RA Tuddenham E.G.D., Mcvey J.H.;
RT "Comparative sequence analysis and molecular evolution of blood
RT coagulation genes from Gallus gallus and Fugu rubripes."
RL Submitted (JAN-2002) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AF465272; AAC33367.1; -
DR GO; GO:0005507; F:Copper ion binding; IEA.
DR GO; GO:0007155; P:cell adhesion; IEA.
DR InterPro; IPR001117; Cu-oxidase.
DR InterPro; IPR008972; Cupredoxin.
DR InterPro; IPR000421; FAS9_C.
DR InterPro; IPR008979; Gal_Bind like.
DR Pfam; PF00394; Cu-oxidase; 1.
DR Pfam; PF00754; F5_F8 type C; 2.
DR SMART; SM00231; FAS8G_2.
DR PROSITE; PS01285; FAS8C_1; 2.
DR PROSITE; PS01286; FAS8C_2; 1.
DR PROSITE; PS00022; FAS8C_3; 2.
DR PROSITE; PS00079; MULTICOPPER_OXIDASE1; 1.
FT NON TER 1
SQ SEQUENCE 1377 AA; 152728 MW; 15ACFD7F765665CD CRC64;
Query Match 20.2%; Score 2511.5; DB 13; Length 1377;
Best local Similarity 36.3%; Pred.No. 5e-141;
Matches 618; Conservative 210; Mismatches 491; Indels 385; Gaps 48;
QY 706 LWIIGCHNSDFRNGMTALLKVVSCDKNTGYDEDSYEDISAVLLSKNAIEPRSPS--Q 763
Db 1 VWTLCGNPFRGRMAKTVTCQCLE-GSLDEEYEDYE----EEDFELQPGFSKRX 55
QY 764 NSRHPSTROKFNATTIPENDIEK-----TDPFAHPTMPKIONVS-SSDLLMLLQSP 817
Db 56 NKRFPVSEQPNNVTS-PKNGTRKPALCLTES--SHRALFHNITNPGSSNGTSAPFGSP 112
QY 818 TPGLSLSLQAEKYE-----TFSDPSPGADDSNNSLSLSEMTHTFRQLH 861
Db 113 HPDDVSTSSLPETTYDVPVSYESFLADEEELSKTISQPGAGALPSRK-----H 160
QY 862 HSGDMVTPESGLQRLNEKLTGTAATELKKLDFKVSSTSNLLSTIPSDNLAAGTNTS 921
Db 161 ISGVRGTVISREGLYQSKAPEDAMVERK-----VINVL 195
QY 922 SLGPPSPMPVHYDSQDLDTTLFGKSSPTESGGPLSLSEENDSKILES-GLMNSQESSWG 980
Db 196 EVQEPFA-----KAAVVQAGGTLQMLETFTQKPMHTYDLMNSTVFAAG 238
QY 981 KNVSTESGRLFKGRAGHPALLTKDNALFKVSISSLKTKTSNNSATNRKTHIDGPELL 1040
Db 239 K-----GPL-----QGTRSSFQDGP--- 254
QY 1041 IENSPSVQNILEDSTPFGKVTPP-IHDMMLMDKNATALRLNHNKNTTSSKNMEMVQOK 1099
Db 255 -EHS-----LGQVTSSEGADVTPLNLHESRPSNTGPTLGSNHNSSFTTDS----- 300
QY 1100 KEGFIPDPAQNPDVSFFKQLFLPSARWIQTHGKNSLNSCGGSPKQVLVLGPEKSVEG 1159
Db 301 --PLGPSARTEDIG-----TSQSHSVESNRSSSELDKLNKRPHKVVSGQFNGSLGG 350
QY 1160 QNFLSEKXKVVVGKEPTKDVGLKEMVPPSSRNFLTNLDMLHNNTNHOEKKTOBE--- 1216
Db 351 ENI-----SFSDLQRVQEQILTDNRNLSLPANSSMEEAAGTFAHNDLEPBYRL 398
QY 1217 -IEKKETLIOENVVLPQIHVTGKNFMKNLFLLLSTRONVEGSDYGAVAPVLQDPRSLND 1275
Db 399 PTERDELLILEAV-----QDQATATKDLPETDLSVLPSQNL----- 434
QY 1276 STNETKKTHTAHFSKKGBEENLEGLNQTKQIVEKYACTRISPNTSQQNFVTOQRKALK 1335
```



```
QY 1996 NLYPGVETVEMLPKAGIWRVECLIGEHLHAGMSTLFLVSNKOOTPLGMASGHIRDFQ 2055
Db 1806 PLIP-----ECPMPGLSTGALSDSQ 1826
QY 2056 ITASGOYGOWAPKLARL--HYSGSIN-AWSTKEPFSWKVDILAPMIHGIKTQGARQKF 2112
Db 1827 IKAS-EY-----LSCIGGHARGLTDPVWYMSFP-DKIPRSFLIPRL-----PKEKY 1871
QY 2113 SLYISQFIMYSLDGKWKQTYGNGTGLMVFNGVDSGIGHNIPNPPIIARYIRLHP 2172
Db 1872 SNAESLSVIIV-----QTTLSLFLVLVYSTAGNSDASTIKENFPDPPIVARYIRIHP 1923
QY 2173 THYSIRSLMELMGCDNSCSMZLGMESKAISDAQITASSYFTNNFAT-WSPSKARLHL 2231
Db 1924 TKSYNRPTELELLGCEVNCSTPLGLEDRIONKOITASSFKKSWGSYWEPSLARLNA 1983
QY 2232 QGSNAWRPOVNNPKWLQVDFQTKWKTGVTTCGVKSLTSMYVKEFLISSQDGHWT 2291
Db 1984 QGRVNAWQAKANNKQWLQIDLLKIKKTAIV-CGCKSLSSSEMYKSYLSYSDQGVSWK 2043
QY 2292 LFTQNGKV--KVFQGNQDSFTPVVNSLDPLLRILRYLRHPQSVVHQIALRMEVLGCEAQD 2349
Db 2044 PYQKSMVDKIFEGNSNTKGHMKNFNPPIISFRILIPKTWNQSLALRIELFGC---D 2100
QY 2350 LY 2351
Db 2101 IY 2102

RESULT 11
Q8BQ43 PRELIMINARY; PRT; 355 AA.
AC Q8BQ43; 2003 (Tremblrel. 23, Created)
DT 01-MAR-2003 (Tremblrel. 23, Last sequence update)
DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)
DE Coagulation factor VIII (Fragment).
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Dorsal root ganglion;
RX MEDLINE=22354683; PubMed=12466853;
RA The FANTOM Consortium,
RA the RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs.";
RL Nature 420:563-573(2002).
DR EMBL; AK051579; BAC34681.1;
DR GO; GO:0005507; F:copper ion binding; IEA.
DR InterPro; IPR001117; Cu-oxidase.
DR Pfam; PF00394; Cu-oxidase; 1.
DR PROSITE; PS00079; MULTICOPPER_OXIDASE1; 1.
FT NON TER 355
SQ SEQUENCE 355 AA; 40548 MW; 2D3C984EFA43F284 CRC64;

Query Match 12.9%; Score 1595.5; DB 11; Length 355;
Best Local Similarity 84.2%; Pred. No. 3.4e-87;
Matches 298; Conservative 22; Mismatches 33; Indels 1; Gaps 1;

QY 1 NQIELSTCFEILCLLFCFSATRRYVLGAVELSWDYMOSD-LGELPVDAREPVRPKSPFF 59
Db 1 NQIALFACFELSLFNLFCFSATRRYVLGAVELSWNVIQSDLLSVLTDGRFPRMSTSPFF 60
QY 60 NTSVYKTLFVEFDHLFNIAKPPPPWMLGPTIOAEVVDTVITLKNMASHPVSLHA 119
Db 61 NTSINVKTVFVEYXQDLFNIAKPPPPWMLGPTIWEVHDTVITLKNMASHPVSLHA 120
QY 120 VGVSYWKASEGAEYDDQTSQREKDDKVPFGSGHTYVWQVLKENGPMASDPLCLTYSYLS 179
Db 120 VGVSYWKASEGAEYDDQTSQREKDDKVPFGSGHTYVWQVLKENGPMASDPLCLTYSYLS 179
```

```
Db 121 VGVSYWKASEGAEYDDQTSQREKDDKVPFGSGHTYVWQVLKENGPMASDPLCLTYSYLS 180
QY 180 HYDLVKDNLNSGLIGALLVCREGSLAKEKTQTLHKFTLLFAVEDEGKSHSETKNSLMQDR 239
Db 181 HYDLVKDNLNSGLIGALLVCREGSLAKEKTQTLHKFTLLFAVEDEGKSHSETKNSLMQDR 240
QY 240 DAASARAWPKHTVAGYVNRSLPGLIGCHRKSVYHVIQMGTTPEVHSTFLSGHTFLVRN 299
Db 241 DSASAEEDWPKMHTVAGYVNRSLPGLIGCHRKSVYHVIQMGTTPEVHSTFLSGHTFLVRN 300
QY 300 HQASLEISPIITFLTAQTLMDLQGLFCHSSHQHDCGAEYVAVKVDSCPEPQ 353
Db 301 HQASLEISPIITFLTAQTLMDLQGLFCHSSHQHDCGAEYVAVKVDSCPEPQ 354

RESULT 12
Q8BQ4X PRELIMINARY; PRT; 745 AA.
AC Q8BQ4X;
DT 01-JUN-2003 (Tremblrel. 24, Created)
DT 01-JUN-2003 (Tremblrel. 24, Last sequence update)
DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)
DE Coagulation factor V precursor (Fragment).
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Aves; Neognathae; Galliformes; Phasianidae; Phasianinae; Gallus.
OX NCBI_TaxID=9031;
RN [1]
SEQUENCE FROM N.A.
RC Davidson C.J., Hirt R.P., Lal K., Snel P., Elgar G.,
RA Tuddenham E.G.D., McVey J.H.;
RT "Comparative sequence analysis and molecular evolution of blood
RT coagulation genes from Gallus gallus and Fugu rubripes.";
RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF465271; AF033661.1;
DR GO; GO:0005507; F:copper ion binding; IEA.
DR GO; GO:0007155; P:cell adhesion; IEA.
DR InterPro; IPR001117; Cu-oxidase.
DR InterPro; IPR008972; Cupredoxin.
DR InterPro; IPR00421; FA58C.
DR InterPro; IPR008973; Gal_bind like.
DR Pfam; PF00394; Cu-oxidase; 1.
DR Pfam; PF00754; F5_F8 tyse_C; 2.
DR SMART; SM00231; FA58C; 2.
DR PROSITE; PS01285; FA58C_1; 2.
DR PROSITE; PS01285; FA58C_2; 2.
DR PROSITE; PS00022; FA58C_3; 2.
DR PROSITE; PS00079; MULTICOPPER_OXIDASE1; 1.
FT NON TER 745
SQ SEQUENCE 745 AA; 85626 MW; A264587A348C29BE CRC64;

Query Match 11.7%; Score 1453; DB 13; Length 745;
Best Local Similarity 39.7%; Pred. No. 3.7e-78;
Matches 299; Conservative 143; Mismatches 256; Indels 54; Gaps 15;

QY 1626 NHAAAINEGONKEIEVTWAKQGRTERLCSQNPVL--KEHQK-----EIRTLQ 1675
Db 9 NHCCVAI--QSHVPSPATENCQGYTQQL---KPSIVGLPHENGNGVEYTSGEYVTDTS 63
QY 1676 SDQEE---IDYDDTISVEMKKEDFDIYDEENQSPRFQK-----KTRHYFTAAVERL 1725
Db 64 GDEYEVYVGFDDPYMTDPK-----LNVNEQRNFDIAEHYLSRGNERYVYAAKEVC 117
QY 1726 WDY-GMSSPHVLNRNAQSGSVPOFKVTVQEFDTGSGFTQPLRGELNEHLGLGPYRA 1784
Db 118 WNYAGLKKSTWYDKIKCKGST---KKVIFQRYTDSFTTVDGEGEEHGLGLGPYRA 174
QY 1785 EVEDNIMVTPRNOASRFYSFYSSLIYSY-----EDQRCGAEPRKNFVKNETKTYFW 1836
Db 175 EVNDVLVHFKNLASRYSLSLHAGLFLYEKSSGSIYDDESSDWFKDDKQVQPNSSYIVW 234
```


QY 705 GLWILGCHNSDFRNRGMTALLKVSSCDKNTGDYEDSYEDISAYLLSKNNNAIPRSPSQN 764
Db 692 GIFEIYQAGSHREEGMAIYVNSQC-----SSHQD-----SPRQHYQA 730
QY 765 SRPSTRQKQFNATTPENDIEKTDWFAHRTMPKIQNVSSDDLMLLRQSTPHGLSL 824
Db 731 SR-----732
QY 825 SDLQAKYETFSDDPSGAISSNNLSGEMTHFRPOLHSGDMVFTPSGGLQALNKLGT 884
Db 733 -----732
QY 885 TAATELKLDFKVSSTGNLSTIPSDNLAAGTNTSSLGPPSMPVHYDSQLDTLFGKY 944
Db 733 -----VY-----735
QY 945 SSPLTESGGPLSSENNDSKLLSGLMNSQBSGKNSVSTBSGRLFKGKRAHGPALLT 1004
Db 736 -----IMAEIHW-----743
QY 1005 KDNALFKVISLTKNTKNTSNAIRKTHIDGPSLLIENSFVSVMONLESDFTEPKVTPL 1064
Db 744 -----DYCPDRSWELE-----WHNTSEKD-----762
QY 1065 IHRMLMXNATALRLNMHGNKTTSSKNMEMVQOKEGPIPPDAQNPDMSFFKVLFPES 1124
Db 763 -----762
QY 1125 ARWIORTHGKNSLNSGQSPKQIOLVSLGPXSVEGONFSLSEKNKVVVGKEFTKDVOLKE 1184
Db 763 -----762
QY 1185 XVFPSSRLFLTNLDNLHNNTHNQEKKIQETEKETLQENVVLQIHTVTGKNFMK 1244
Db 763 -----SYGHVFLSKD-----773
QY 1245 NLFLSTRQNVBGSYGAYAPVLQDRSLNDSTNRKHTAHFSSKKEENLEGLGNQTK 1304
Db 774 -----773
QY 1305 QIVEKYACTFRISNTSQNFVORSKALKQFLPLLEETELEKEIIVDDTSTOWSKMK 1364
Db 774 -----773
QY 1365 HLTPSTLTQIDYNEKKEGKAITQSPSLDCLTRSHSIPQANRSPPIAKVSPSPSRPIVLT 1424
Db 774 -----773
QY 1425 RVLPQDNSSHLPAASYRKKSQGVQESSHFLQGAKKNNLSLAILTLEMTGDQREVSGSLGTS 1484
Db 774 -----GLLGS-----778
QY 1485 ATNSVTYKVENTVLPKPDLEKTSKVELLPKVHLYOKDLFPTSTNSGSPGHLDLVEGSL 1544
Db 779 -----778
QY 1545 LQGTGAIKWNEANRPGKVPFLRVATESSAKTPSKLLDPLANDNEYGTQIPKEBKWSQEK 1604
Db 779 -----778
QY 1605 SPEKTAFKKDTILSLNACENHAIANECQKPELEVTWAKQGRTERLCSQNPVULKR 1664
Db 779 -----778
QY 1665 HOREITRTLQSDQBEIDYDDTISVMKKEDFDIVDEDENSPSFQKKTRHYFLAAVER 1724
Db 779 -----778
QY 1725 LWDYGMSSSPHLNRNAQSGVPOFKVVPQBFDTGSPTOPLRYGELNEHILGLLGPYIRA 1784
Db 779 -----KYKKAFFREYTDGTRIPRPSGPEEHLGILGPLIRG 815
QY 1785 EVDNIMVTFRNQASRPYSFYS-SLISYEEDQROGAEPKPNKPNETKTYFNKVQHMA 1843

Db 816 EVGDILTIVVFKKASRPYSTHAHGVLESNTGGQAAE-----PGEVLTYQWNIPIERSG 868
QY 1844 PTKDEFDCXAWAYFSDVLEKDVHSGJLGLPLVCHTINTLNPAHGRQVTVQEPALFTIPD 1903
Db 869 PGSDSACYSWIYSAVDPIKDMYSGLVGLVTCRNGILPEPGRNDMDREFALLFLIFD 928
QY 1904 ETKSWYFTENMER-NCRAPCNIQMEDPTFKENYRFAHNGYIMDTLFGVMAQDQIRWY 1962
Db 929 ENQSWLKNINAYGFCSSHVNLKDATFLESKNHAINGLKLYANLRLGLTVYGERVAVY 988
QY 1963 LLSNGNENIHSFHSGHVTVRKBEYKVALNLYPGVFETVEMLPFSKAGIWRVECLIG 2022
Db 989 MLANGQDTHIVHFAESPLYQNGOSYRADVVDLPFGTPEVVEMVASNEGTVLMCHCHVT 1048
QY 2023 EHLHAGMSTLFLVYSNK 2039
Db 1049 DHVHAGMETIFTVLISHE 1065

RESULT 14
Q9Z0Z4
ID Q9Z0Z4 PRELIMINARY; PRT; 1157 AA.
AC Q9Z0Z4; 1999 (Tremblrel. 10, Created)
DT 01-MAY-1999 (Tremblrel. 10, Last sequence update)
DT 01-MAY-1999 (Tremblrel. 10, Last sequence update)
DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)
DE Hephaestin.
GN HSPH.
OS Mus musculus (Mouse).
OC Bukayota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J;
RX MEDLINE=99140771; PubMed=9988272;
RA Valpe C.D., Kuo Y.M., Murphy T.L., Cowley L., Askwith C., Libina N.,
RA Gitschier J., Anderson G.J.;
RT "Hephaestin, a ceruloplasmin homologue implicated in intestinal iron
transport, is defective in the sla mouse."
RL Nat. Genet. 21:195-199(1999).
DR EMBL; AF082567; AADI6035.1; -
DR HSP; P00450; IKCW.
DR MGD; MGI:1332240; Heph.
DR GO; GO:0005507; F:copper ion binding; IEA.
DR InterPro; IPR001117; Cu-oxidase.
DR InterPro; IPR008972; Cupredoxin.
DR Pfam; PF00394; Cu-oxidase; 3.
DR PROSITE; PS00079; MULTICOPPER OXIDASE1; 3.
DR PROSITE; PS00080; MULTICOPPER OXIDASE2; 1.
SQ SEQUENCE 1157 AA; 129682 MW; 411BBIDA2BFEC2PD CRC64;

Query Match 11.2%; Score 1389; DB 11; Length 1157;
Best local Similarity 20.0%; Pred. No. 4.9e-74;
Matches 412; Conservative 179; Mismatches 414; Indels 1052; Gaps 26;
QY 20 ATREYVILGAVELSDYVMSQSDIGLPELVDARPPRPVKSF----PENTSVVYKKTFLVFEFTD 75
Db 24 AIRNYVIGIQDMQWNYAPKGRNVITNQTLLNDTVASSFLSKGNRISSYKTVYKEYSD 83
QY 76 HLFNIAKPRPFWMLGLGTIOAEVYDVTVTILKNMASHPVSLHAGVSYWKAEGAYDD 135
Db 84 GTYTEETAKPAWLGLPLQLQAEVGDVILHLKNAFASRPVTHPHGVFEKDSGSLYPD 143
QY 136 QTSOREKDDKVPKSGSHTYVQVLEKNGPMASDPLCLITYSLSHVDLVLDLNSGLIGAL 195
Db 144 GSSGLKADSDVPPGGSHVYVNSIPESHAPTEADPACLTWIYHSHVDAPRDIATLGLPL 203
QY 196 LVCREGSI----AKEQTQLHKFTLLFAVDEGKSWHE-----TKNSLMQDRDAASAR 245
Db 204 ITCRGLDGNSPPRKDVNDHNFLLFSVIDENLSWHLDDNIATYCDPASPVDKEDGAFQ 263

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 13, 2004, 14:01:28 ; Search time 22.5 Seconds
(without alignments)
5394.340 Million cell updates/sec

Title: NP000123-328-355-581

Perfect score: 12415

Sequence: 1 MQELSTCFLLRPFCSA.....WFOQLRMEVLGCEAQQDLY 2351

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA:*

1: /cgn2_6/ptodata/2/iaa/5A COMB.pap:*

2: /cgn2_6/ptodata/2/iaa/5B COMB.pap:*

3: /cgn2_6/ptodata/2/iaa/6A COMB.pap:*

4: /cgn2_6/ptodata/2/iaa/6B COMB.pap:*

5: /cgn2_6/ptodata/2/iaa/PCPUS COMB.pap:*

6: /cgn2_6/ptodata/2/iaa/backfiles1.pap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	12399	99.9	2351	4	Sequence 4, Appli
2	12393	99.8	2351	6	Patent No. 5171844
3	12390	99.8	2351	1	Sequence 2, Appli
4	12386	99.8	2351	1	Sequence 2, Appli
5	12377	99.7	2351	6	Patent No. 5422260
6	12282	98.9	2332	1	Sequence 4, Appli
7	12282	98.9	2332	1	Sequence 2, Appli
8	12282	98.9	2332	1	Sequence 2, Appli
9	12282	98.9	2332	1	Sequence 2, Appli
10	12282	98.9	2332	1	Sequence 2, Appli
11	12282	98.9	2332	3	Sequence 2, Appli
12	12282	98.9	2332	3	Sequence 2, Appli
13	12282	98.9	2332	4	Sequence 2, Appli
14	12282	98.9	2332	5	Sequence 2, Appli
15	12282	98.9	2332	5	Sequence 2, Appli
16	12275	98.9	2332	1	Sequence 2, Appli
17	12232	98.5	2332	3	Sequence 2, Appli
18	9474	76.3	2343	3	Sequence 2, Appli
19	8820	71.0	2319	1	Sequence 2, Appli
20	8820	71.0	2319	1	Sequence 2, Appli
21	8820	71.0	2319	3	Sequence 2, Appli
22	8820	71.0	2319	3	Sequence 2, Appli
23	8820	71.0	2319	4	Sequence 2, Appli
24	8820	71.0	2319	4	Sequence 2, Appli
25	8820	71.0	2319	5	Sequence 2, Appli
26	8715	70.2	2304	3	Sequence 2, Appli
27	8338	67.2	1661	2	Sequence 2, Appli

28	8338	67.2	1661	2	US-08-358-107-2	Sequence 2, Appli
29	8338	67.2	1661	2	US-09-243-539-2	Sequence 2, Appli
30	8057	64.9	2133	3	US-08-670-707A-37	Sequence 37, Appli
31	8057	64.9	2133	3	US-09-037-601-37	Sequence 37, Appli
32	8057	64.9	2133	4	US-09-315-179-37	Sequence 30, Appli
33	8051	64.8	2133	4	US-09-523-656-30	Sequence 30, Appli
34	7940.5	64.0	2115	3	US-09-324-867-5	Sequence 5, Appli
35	7381	59.5	1471	1	US-08-683-839B-3	Sequence 3, Appli
36	7208	58.1	1438	4	US-09-209-916-1	Sequence 1, Appli
37	6205	50.0	1467	4	US-09-523-656-38	Sequence 38, Appli
38	6103	49.2	1443	2	US-08-670-707A-39	Sequence 39, Appli
39	6103	49.2	1443	3	US-09-037-601-39	Sequence 39, Appli
40	6103	49.2	1443	3	US-09-315-179-39	Sequence 39, Appli
41	2771.5	22.3	2183	3	US-08-746-111-5	Sequence 5, Appli
42	2425.5	19.5	541	1	US-08-121-202-4	Sequence 4, Appli
43	2258	18.2	439	3	US-08-448-722A-5	Sequence 5, Appli
44	2258	18.2	439	3	US-08-189-309B-5	Sequence 5, Appli
45	2130.5	17.6	868	1	US-07-864-004B-6	Sequence 6, Appli

ALIGNMENTS

RESULT 1

US-10-133-907-4

; Sequence 4, Application US/10133907

; Patent No. 6677369

; GENERAL INFORMATION:

; APPLICANT: Chien, Kenneth R

; TITLE OF INVENTION: Method to treat hemophilia by hepatic gene transfer of Factor VIII

; FILE REFERENCE: 6627-P1170

; CURRENT APPLICATION NUMBER: US/10/133,907

; PRIOR FILING DATE: 2002-04-25

; PRIOR APPLICATION NUMBER: 60/286,314

; NUMBER OF SEQ ID NOS: 5

; SOFTWARE: PatentIn version 3.1

; SEQ ID NO 4

; LENGTH: 2351

; TYPE: PRT

; ORGANISM: Homo sapiens

; US-10-133-907-4

Query Match	98.9%	Score	12399	DB	4	Length	2351
Best Local Similarity	99.9%	Pred. No.	0				
Matches	2348	Conservative	1	Mismatches	2	Indels	0
Gaps	0						
Qy	1	MQELSTCFLLRPFCSA	TRRYLGA	VELSWDYM	QSDLGELFVDARFPPR	PKSPFN	60
Db	1	MQELSTCFLLRPFCSA	TRRYLGA	VELSWDYM	QSDLGELFVDARFPPR	PKSPFN	60
Qy	61	TSVYKTKLFEVFDHLFNI	AKRPPM	WGLLGPTIQAEVDTVVIT	LKNMASHPVSHAV	120	
Db	61	TSVYKTKLFEVFDHLFNI	AKRPPM	WGLLGPTIQAEVDTVVIT	LKNMASHPVSHAV	120	
Qy	121	GVSWKASEGAEXDDOTS	OREKED	KVPGSSH	YVMQVLKENG	PMASDPLCLTYSYL	180
Db	121	GVSWKASEGAEXDDOTS	OREKED	KVPGSSH	YVMQVLKENG	PMASDPLCLTYSYL	180
Qy	181	VDLVKDLNSGLIGALLV	CREGSLAK	ETQTLHK	FILLFAV	DEGKSWHSEKNSLMQ	240
Db	181	VDLVKDLNSGLIGALLV	CREGSLAK	ETQTLHK	FILLFAV	DEGKSWHSEKNSLMQ	240
Qy	241	AASARAPKCHTNGVY	NRSPLG	IGCHRSVY	WHVIGM	GTTPVHSIFLEGTFLVR	300
Db	241	AASARAPKCHTNGVY	NRSPLG	IGCHRSVY	WHVIGM	GTTPVHSIFLEGTFLVR	300
Qy	301	RQASLEISPTITLTA	OTLLMD	LQGLF	LSCHISS	HOHDMAYKVDSCPEE	360
Db	301	RQASLEISPTITLTA	OTLLMD	LQGLF	LSCHISS	HOHDMAYKVDSCPEE	360

361 EAEDYDDDLTSEMDVAFDDDNPSFQIRSAVAKKHPTWVHYIAAEEEDWDVAPLV.A 420
361 EABYDDDLTSEMDVAFDDDNPSFQIRSAVAKKHPTWVHYIAAEEEDWDVAPLV.A 420
421 PDDRSYKSOYLNNQPORIGRKYKVRWAYTDFKTREAQHSGSILGLPLYGEVGTLL 480
421 PDDRSYKSOYLNNQPORIGRKYKVRWAYTDFKTREAQHSGSILGLPLYGEVGTLL 480
481 LIIFKQASRYNIPYHGHITDVRPLYSRLPKGVKHLKDPILPGEIIFYKWTVTVEDGP 540
481 LIIFKQASRYNIPYHGHITDVRPLYSRLPKGVKHLKDPILPGEIIFYKWTVTVEDGP 540
541 TKSDPRCLTRYSSFWMMERDLASGLIPLICYKESVDQGNQIMSKRNWILFVSFDE 600
541 TKSDPRCLTRYSSFWMMERDLASGLIPLICYKESVDQGNQIMSKRNWILFVSFDE 600
601 NRSWLTNINORPLNPAGVOLDEDFQASNTMHSGINGYVFDLSQLSVCLHEVAVWYILS 660
601 NRSWLTNINORPLNPAGVOLDEDFQASNTMHSGINGYVFDLSQLSVCLHEVAVWYILS 660
661 IGAOTDFLSVPSGVTFKHKMVRDRTLPLPPSGETVFSMENPGLWILGCHNSDFNRG 720
661 IGAOTDFLSVPSGVTFKHKMVRDRTLPLPPSGETVFSMENPGLWILGCHNSDFNRG 720
721 MTALLKVSCKNTGDYVEDSVEDISAYLLSKNNAIEPRSFQNSRHPSTRQKQFNATTI 780
721 MTALLKVSCKNTGDYVEDSVEDISAYLLSKNNAIEPRSFQNSRHPSTRQKQFNATTI 780
781 PENDIEKTDPAHPTMPKTONVSSDLMLLRQSPTHGLSLSDLOEAKYEFSDDPSS 840
781 PENDIEKTDPAHPTMPKTONVSSDLMLLRQSPTHGLSLSDLOEAKYEFSDDPSS 840
841 PCAIDSNLSMTHFRQLHSHGDMVTPESGLQRLNEKLGTTAAATELKLDKFKVST 900
841 PCAIDSNLSMTHFRQLHSHGDMVTPESGLQRLNEKLGTTAAATELKLDKFKVST 900
901 SNLLSTIPSDNLAAGTNTSSLGPPSPVHYDSQDITFLFKKSSPLTESGPPLSLSEE 960
901 SNLLSTIPSDNLAAGTNTSSLGPPSPVHYDSQDITFLFKKSSPLTESGPPLSLSEE 960
961 NNDKSLISGLMNSOESGKNVSTESGRIFKGRAGFALLTKDNALFKVSISSLKTN 1020
961 NNDKSLISGLMNSOESGKNVSTESGRIFKGRAGFALLTKDNALFKVSISSLKTN 1020
1021 KTSNNSATNRKTHIDGPSLLIENSFSVMQNILESUTEFKKVTPLIHDRMLMDKNATLRL 1080
1021 KTSNNSATNRKTHIDGPSLLIENSFSVMQNILESUTEFKKVTPLIHDRMLMDKNATLRL 1080
1081 NMSNKTSSKNMVMVQKKKEGPIPDQONPDMSPFKMLFLPESARWQRTHGKNSLNSG 1140
1081 NMSNKTSSKNMVMVQKKKEGPIPDQONPDMSPFKMLFLPESARWQRTHGKNSLNSG 1140
1141 CGPSFKQVSLGPEKSVRGONFLSKNKVVCGKEFTKDVGLKEMVFPSSRNLFUTNLDN 1200
1141 CGPSFKQVSLGPEKSVRGONFLSKNKVVCGKEFTKDVGLKEMVFPSSRNLFUTNLDN 1200
1201 LHENNTNHOEKIIOBEIEKKEKTELIOENVVLPOHVTGTCTKPFMKXLLFLSTQNVEGSD 1260
1201 LHENNTNHOEKIIOBEIEKKEKTELIOENVVLPOHVTGTCTKPFMKXLLFLSTQNVEGSD 1260
1261 GAYAPVLQDFRSLNDSNTRKHTAHFSSKGBEENLEGNGQTKQIVERYACTTRISNPT 1320
1261 GAYAPVLQDFRSLNDSNTRKHTAHFSSKGBEENLEGNGQTKQIVERYACTTRISNPT 1320
1321 SOQNFTQSRKALQFRLPLETELEKELIIVDDTSTOWSKMKHLTPSTLQIDYNEKE 1380
1321 SOQNFTQSRKALQFRLPLETELEKELIIVDDTSTOWSKMKHLTPSTLQIDYNEKE 1380
1381 KGAIQTOSPLSDCLTRSHSIFQANRGPPIAKVSSPSPRIPIYLRVLFQDNSSHLPAAAY 1440
1381 KGAIQTOSPLSDCLTRSHSIFQANRGPPIAKVSSPSPRIPIYLRVLFQDNSSHLPAAAY 1440
1441 RKKDSGOVQESSHPTQGAKKNNLSLAILTLEMTQDQREVGSLGTSATNSVTYKVENVTLP 1500

1441 RKKDSGOVQESSHPTQGAKKNNLSLAILTLEMTQDQREVGSLGTSATNSVTYKVENVTLP 1500
1501 KPDLPKTSGKVELLPKVHIYQKOLPPTETSNNGSPGHLDLVEGSLLOGTEGAIKWNEANRP 1560
1501 KPDLPKTSGKVELLPKVHIYQKOLPPTETSNNGSPGHLDLVEGSLLOGTEGAIKWNEANRP 1560
1561 GKVFELAVATESSAKTPSKLLDPLAWNHYGTQIPKBEWKSQESPKTAPKPKDXTILSL 1620
1561 GKVFELAVATESSAKTPSKLLDPLAWNHYGTQIPKBEWKSQESPKTAPKPKDXTILSL 1620
1621 NACESNHAIAINEGQNKPEIEVTWAKQGRTERLCSQNPVVLKHHQREBITRTTILQSDQEE 1680
1621 NACESNHAIAINEGQNKPEIEVTWAKQGRTERLCSQNPVVLKHHQREBITRTTILQSDQEE 1680
1681 IDYDDTISVEMKK3DFDIYDEDENOSPRSPOKKTRHYFLAAVERLWDYGMSSSSHVLNRR 1740
1681 IDYDDTISVEMKK3DFDIYDEDENOSPRSPOKKTRHYFLAAVERLWDYGMSSSSHVLNRR 1740
1741 AQS5VPOFKKVPQEFETDGSFTQPLVRGELNEHLGLLGPYIRAEVEDNIMVTFRNOASR 1800
1741 AQS5VPOFKKVPQEFETDGSFTQPLVRGELNEHLGLLGPYIRAEVEDNIMVTFRNOASR 1800
1801 PYSYSSLIISYEEDORQAGPRKPNFVKNPTKTYFWKQHHMAPTKDFDCAWAYPSDV 1860
1801 PYSYSSLIISYEEDORQAGPRKPNFVKNPTKTYFWKQHHMAPTKDFDCAWAYPSDV 1860
1861 DLEKDVHSLTGILLVCHTNTLNPAHGRONTVOEFALPFTTIFDETKSWYFTENVERNCRA 1920
1861 DLEKDVHSLTGILLVCHTNTLNPAHGRONTVOEFALPFTTIFDETKSWYFTENVERNCRA 1920
1921 PCNIQMEDPTFKENYRPHAINGYIMDTLPGIWMADQORIRWYLLSMGSNENIHSIFSGH 1980
1921 PCNIQMEDPTFKENYRPHAINGYIMDTLPGIWMADQORIRWYLLSMGSNENIHSIFSGH 1980
1981 VFTVRKKEEYKMALYNLYPGVFETVEMLP3KAGIWRVBCCLIGEHLHAGMSTLFLVYSNKC 2040
1981 VFTVRKKEEYKMALYNLYPGVFETVEMLP3KAGIWRVBCCLIGEHLHAGMSTLFLVYSNKC 2040
2041 QTPGLMAGSHIRDFOITASGOYGOWAPKLABLHYSGSINAWSTKEPFSWIKVDLAPMII 2100
2041 QTPGLMAGSHIRDFOITASGOYGOWAPKLABLHYSGSINAWSTKEPFSWIKVDLAPMII 2100
2101 HGIKTQGARQKFSLLYISQFIIMTSLDGKKQWTVRGNSTGTLMVFFGNVDSGSKHNIEN 2160
2101 HGIKTQGARQKFSLLYISQFIIMTSLDGKKQWTVRGNSTGTLMVFFGNVDSGSKHNIEN 2160
2161 PPIIARYIRLPHTHYSIRSTRMLMGLDNLSCMPLGMEKSAISDAQITASSYFTNMFA 2220
2161 PPIIARYIRLPHTHYSIRSTRMLMGLDNLSCMPLGMEKSAISDAQITASSYFTNMFA 2220
2221 TWSPSKARLHLQGRSNAWRPOVNNPKEWLQVDFOKTKVGTGVTQGVKSHLTSYVYKBEFL 2280
2221 TWSPSKARLHLQGRSNAWRPOVNNPKEWLQVDFOKTKVGTGVTQGVKSHLTSYVYKBEFL 2280
2281 ISSQDGHQWTLFTQNGKVKVQGNQDSFTPVNSLDPPLLTRYLRIHPQSWHQAIALRM 2340
2281 ISSQDGHQWTLFTQNGKVKVQGNQDSFTPVNSLDPPLLTRYLRIHPQSWHQAIALRM 2340
2341 EVLGCQEAQDLY 2351
2341 EVLGCQEAQDLY 2351

RESULT 2

5171844-2

; Patent No. 5171844

; APPLICANT: VAN COYEN, ALBERT J. J.; PANNEKOEK, HANS; VERBEEET,

; MARTINUS P.; VAN LEEEN, ROBERT W.

; TITLE OF INVENTION: PROTEINS WITH FACTOR VIII ACTIVITY

; PROCESS FOR THEIR PREPARATION USING GENETICALLY-ENGINEERED CELLS

; AND PHARMACEUTICAL COMPOSITIONS CONTAINING THEM

; NUMBER OF SEQUENCES: 12


```

; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/205.226
; FILING DATE: 10-JUN-1988
; SEQ ID NO:2
; LENGTH: 2351
5171644-2

Query Match      99.8%; Score 12393; DB 6; Length 2351;
Best local Similarity 99.8%; Pred. No. 0;
Matches 2347; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy 1 MQIELSCFFFLCLLRFCSATRRYYLGAVALUSWDYMSDLGELVDARFPPRVKSPFEN 60
Db 1 MQIELSCFFFLCLLRFCSATRRYYLGAVALUSWDYMSDLGELVDARFPPRVKSPFEN 60

Qy 61 TSVVYKXKL FVEFTDHLFNIAKPPPPWGLLGP--TQAEVYDVTVITLKNMASHVPSLEAV 120
Db 61 TSVVYKXKL FVEFTDHLFNIAKPPPPWGLLGP--TQAEVYDVTVITLKNMASHVPSLEAV 120

Qy 121 GVSWKASGAEYDCQTSOREKEDDKVPFGSGHTYVWQVILKENGPMASDPLCLTYSLSH 180
Db 121 GVSWKASGAEYDCQTSOREKEDDKVPFGSGHTYVWQVILKENGPMASDPLCLTYSLSH 180

Qy 121 GVSWKASGAEYDCQTSOREKEDDKVPFGSGHTYVWQVILKENGPMASDPLCLTYSLSH 180
Db 121 GVSWKASGAEYDCQTSOREKEDDKVPFGSGHTYVWQVILKENGPMASDPLCLTYSLSH 180

Qy 161 VDLVKDLSGLIGALLVCREGSLAKEKQTOLHKELLFAVDEGKSWHSETKNSLMQDRD 240
Db 161 VDLVKDLSGLIGALLVCREGSLAKEKQTOLHKELLFAVDEGKSWHSETKNSLMQDRD 240

Qy 241 AASARAPKMTVNGVYVNRSLPGJGCHRXKSVWHVIGMGTTPVHSTFLEGHFTLVNRH 300
Db 241 AASARAPKMTVNGVYVNRSLPGJGCHRXKSVWHVIGMGTTPVHSTFLEGHFTLVNRH 300

Qy 301 ROASLEISPIITFLTAQTLMDLGOPLLSCHLSSHOHDGMEAYVVDSCPEPQLRMKNE 360
Db 301 ROASLEISPIITFLTAQTLMDLGOPLLSCHLSSHOHDGMEAYVVDSCPEPQLRMKNE 360

Qy 361 EAEDYDDLTDSEMDVRFDDNDSFQIRSVAKKHPTKVHYIAAEEEDWDVAPLVLA 420
Db 361 EAEDYDDLTDSEMDVRFDDNDSFQIRSVAKKHPTKVHYIAAEEEDWDVAPLVLA 420

Qy 421 PDDRSYKSOYLNGPQRIRGRYKXKVRPMAYTDETPKTBREAIOHESGILGPLLYGEGVDTL 480
Db 421 PDDRSYKSOYLNGPQRIRGRYKXKVRPMAYTDETPKTBREAIOHESGILGPLLYGEGVDTL 480

Qy 481 LIIFKNOQASRPVNIYPHGITDVRPIYSRRLPKGVKHLKDFPLPGEIKYKWTVTVEDGP 540
Db 481 LIIFKNOQASRPVNIYPHGITDVRPIYSRRLPKGVKHLKDFPLPGEIKYKWTVTVEDGP 540

Qy 541 TKSDPRCLTRYYSFVNMRDLASGLIGPLLYCYKESVDQKGNQMSDKRNVLFSVFE 600
Db 541 TKSDPRCLTRYYSFVNMRDLASGLIGPLLYCYKESVDQKGNQMSDKRNVLFSVFE 600

Qy 601 NRSWYLTENIQRFLENPAGVQLEDPEFOASNIMHSINGVYVDSLOLSVCLHEVAYWYILS 660
Db 601 NRSWYLTENIQRFLENPAGVQLEDPEFOASNIMHSINGVYVDSLOLSVCLHEVAYWYILS 660

Qy 661 IGAQTDLSVPESGVTEKHMYVEDTLTFPFSGETVFMSPENGLWILGCHNSFRNRG 720
Db 661 IGAQTDLSVPESGVTEKHMYVEDTLTFPFSGETVFMSPENGLWILGCHNSFRNRG 720

Qy 721 MTALLKVSQCDKNTGDYVSDYEDISAYLLSKNNAIEPRSFQNSRHPSTROKQFNATTI 780
Db 721 MTALLKVSQCDKNTGDYVSDYEDISAYLLSKNNAIEPRSFQNSRHPSTROKQFNATTI 780

Qy 781 PENDTEKTDPTFAHRTPEMKI QNVSSDILLMLRQSPHGLSLSDLOZAKYETETSDDPS 840
Db 781 PENDTEKTDPTFAHRTPEMKI QNVSSDILLMLRQSPHGLSLSDLOZAKYETETSDDPS 840

Qy 841 PGAIISNNLSLSEWTHRPOLHSGDMVFTPEGLQLRLNEKLGTGTAATELKKLDEKVSST 900
Db 841 PGAIISNNLSLSEWTHRPOLHSGDMVFTPEGLQLRLNEKLGTGTAATELKKLDEKVSST 900

Qy 901 SNNLSTIPSDNLAAGTNTSSIGPPSPMPVHYDQJLDTTLFGKKSPLTESGPGSLSEE 960

```

```

Db 901 SNNLSTIPSDNLAAGTNTSSIGPPSPMPVHYDQJLDTTLFGKKSPLTESGPGSLSEE 960
Qy 961 NNDSKLLESGIAMSQSSWGKQVSVSTESGBLFGKGAHGPAALLTKDNALPKVSIILLKTN 1020
Db 961 NNDSKLLESGIAMSQSSWGKQVSVSTESGBLFGKGAHGPAALLTKDNALPKVSIILLKTN 1020
Qy 1021 KTSNNSATNRKTHIDGSPSLLEIENSPSVWQNLLESDETFKAVTPLIHDRMLMDKNATLRL 1080
Db 1021 KTSNNSATNRKTHIDGSPSLLEIENSPSVWQNLLESDETFKAVTPLIHDRMLMDKNATLRL 1080
Qy 1081 NMSNKTTSNNMEMVQCKXEGFIPDQAQPDMSFFKMLFLPESARWIQTHGKNSLNSG 1140
Db 1081 NMSNKTTSNNMEMVQCKXEGFIPDQAQPDMSFFKMLFLPESARWIQTHGKNSLNSG 1140
Qy 1141 QGSPKQIVSLGPEKSVGEGQNFLEKKNVVGGEPTKQVGLKEMVFPSSRNLFITLND 1200
Db 1141 QGSPKQIVSLGPEKSVGEGQNFLEKKNVVGGEPTKQVGLKEMVFPSSRNLFITLND 1200
Qy 1201 LHENNTNQEKKIQEIEKKEETLQENNVLPQIHTVTGTKNFMKNLFLLSRQNVESYD 1260
Db 1201 LHENNTNQEKKIQEIEKKEETLQENNVLPQIHTVTGTKNFMKNLFLLSRQNVESYD 1260
Qy 1261 GAYAPVLQDPRSLNDSNRTRKHTAHFSPKKEEENRGLGNQTKQIVKACATTRISPT 1320
Db 1261 GAYAPVLQDPRSLNDSNRTRKHTAHFSPKKEEENRGLGNQTKQIVKACATTRISPT 1320
Qy 1321 SOQNFTVQSRKRALKQRLPLEETELEKRIIVDDTSTQWKNMKHLTPESTLQIDYNEKE 1380
Db 1321 SOQNFTVQSRKRALKQRLPLEETELEKRIIVDDTSTQWKNMKHLTPESTLQIDYNEKE 1380
Qy 1381 KGATITQSPSLDCLTRSHSIPQANRSPPIAKVSPSPSIRPIYLRVLFDQNSSHELPAASY 1440
Db 1381 KGATITQSPSLDCLTRSHSIPQANRSPPIAKVSPSPSIRPIYLRVLFDQNSSHELPAASY 1440
Qy 1441 RKXDSGVQESSEFQOGAKNNLSLAITLMTGQREVSGTSGATNSVTYKQVNTVLP 1500
Db 1441 RKXDSGVQESSEFQOGAKNNLSLAITLMTGQREVSGTSGATNSVTYKQVNTVLP 1500
Qy 1501 KPDLPKTSGKVELLPKHVYQKDLFPPTETSNQSGHLDLVEGSLLOGTEGAIKNEANRP 1560
Db 1501 KPDLPKTSGKVELLPKHVYQKDLFPPTETSNQSGHLDLVEGSLLOGTEGAIKNEANRP 1560
Qy 1561 GKVPFLRVATESSAKTPSKLLDPLAWDNHYGTQPKBEWKSQESKPEKTKDITLSL 1620
Db 1561 GKVPFLRVATESSAKTPSKLLDPLAWDNHYGTQPKBEWKSQESKPEKTKDITLSL 1620
Qy 1621 NACSNHAAIAINGONKQKPEIEVTWAKQGRTERLCSQNPVVLKHHORBITTTLOSQDEE 1680
Db 1621 NACSNHAAIAIAINEGONKQKPEIEVTWAKQGRTERLCSQNPVVLKHHORBITTTLOSQDEE 1680
Qy 1681 IDYDDTISVEKKKDFDIYDEDENQSPRSFQKTRHYFIAAVERLDYGMSSSPHVLNR 1740
Db 1681 IDYDDTISVEKKKDFDIYDEDENQSPRSFQKTRHYFIAAVERLDYGMSSSPHVLNR 1740
Qy 1741 AQSGSVPOFKVVFQEFDTGSGFTQPLRYGELNEHLGILGPYIRAEVEDNIMVTFRNOASR 1800
Db 1741 AQSGSVPOFKVVFQEFDTGSGFTQPLRYGELNEHLGILGPYIRAEVEDNIMVTFRNOASR 1800
Qy 1801 PYSYSSLSIYEEBQORQABRPKPFVZNEKTYFWKQHHMAKDFDCKAWAYFSDV 1860
Db 1801 PYSYSSLSIYEEBQORQABRPKPFVZNEKTYFWKQHHMAKDFDCKAWAYFSDV 1860
Qy 1861 DLEKDVHSLGILPLLVCHTNTLNPAHGRQVTVQBFALPFTTFDETKSWYFTENMERNCRA 1920
Db 1861 DLEKDVHSLGILPLLVCHTNTLNPAHGRQVTVQBFALPFTTFDETKSWYFTENMERNCRA 1920
Qy 1921 PCNIQMEDPTFKENYRPHAINGYIMDTLPGLVMAQDQQRIRWYLLSMGNSNENIHSIHFGH 1980
Db 1921 PCNIQMEDPTFKENYRPHAINGYIMDTLPGLVMAQDQQRIRWYLLSMGNSNENIHSIHFGH 1980
Qy 1981 VPTVRK3EYKMALYNLYPGVPEVTEMLPSKAGIWRVECLIGEHLAQMSTLFLVYSNKC 2040
Db 1981 VPTVRK3EYKMALYNLYPGVPEVTEMLPSKAGIWRVECLIGEHLAQMSTLFLVYSNKC 2040

```

QY 2041 QTPLOMASCHIRDFQITASGOYQOWAPKLABLHSGSINASTKEPFSWIKVDLLAPMII 2100
DB 2041 QTPLOMASCHIRDFQITASGOYQOWAPKLABLHSGSINASTKEPFSWIKVDLLAPMII 2100
QY 2101 HGKKTQARQKFSSLIYISQFIIMYSLDGKKWQYRGNSGTGILMVFPGNVDSSGKHNIFN 2160
DB 2101 HGKKTQARQKFSSLIYISQFIIMYSLDGKKWQYRGNSGTGILMVFPGNVDSSGKHNIFN 2160
QY 2161 PPIIARYIRLHPHTYISIRSTLRMLMGCDJNSCSMPLGMSKASIDAQITASSYFTNMFA 2220
DB 2161 PPIIARYIRLHPHTYISIRSTLRMLMGCDJNSCSMPLGMSKASIDAQITASSYFTNMFA 2220
QY 2221 TWSPEKARHLQGRSNARPOVNNPEKELQVDFQKTMKVTGVTGQVKSLLTSMYVKEFL 2280
DB 2221 TWSPEKARHLQGRSNARPOVNNPEKELQVDFQKTMKVTGVTGQVKSLLTSMYVKEFL 2280
QY 2281 TSSSDGCHQWTLFFQNGKVKVQGNQDFTPVVNSLDPLLTTRYLRIRHPQSWHQAIALRM 2340
DB 2281 TSSSDGCHQWTLFFQNGKVKVQGNQDFTPVVNSLDPLLTTRYLRIRHPQSWHQAIALRM 2340
QY 2341 EVLGCEAQDLY 2351
DB 2341 EVLGCEAQDLY 2351

RESULT 3

US-08-366-851A-2
; Sequence 2, Application US/08366851A
; Patent No. 5681746
; GENERAL INFORMATION:
; APPLICANT: Bodner, Mordechai
; APPLICANT: De Polo, Nicolas J.
; APPLICANT: Hsu, David Chi-Pang
; APPLICANT: Chang, Steven
; TITLE OF INVENTION: Retroviral Delivery of Full Length Factor VIII
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Viagene, Inc.
; STREET: 11055 Roselle Street
; CITY: San Diego
; STATE: California
; COUNTRY: U.S.A.
; ZIP: 92121
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/366,851A
; FILING DATE:
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Chambers, Daniel M.
; REGISTRATION NUMBER: 34,561
; REFERENCE/DOCKET NUMBER: 930049.438
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 452-1288
; TELEFAX: (619) 452-2616
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2351 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-366-851A-2
Query Match 99.8%; Score 12390; DB 1; Length 2351;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 2347; Conservative 1; Mismatches 3; Indels 0; Gaps 3;
QY 1 MQIELSTCFELCLLRFCFSATRRYYLGAVELSDWYMQSDLGELPVDARFPFRPKSFPFN 60

DB 1 MQIELSTCFELCLLRFCFSATRRYYLGAVELSDWYMQSDLGELPVDARFPFRPKSFPFN 60
QY 61 TSVVYKKTFLFVEFTDHLFNIAKPRPPMGLGPTIQAEVYDVTVVITLKNMASHPSVSHAV 120
DB 61 TSVVYKKTFLFVEFTDHLFNIAKPRPPMGLGPTIQAEVYDVTVVITLKNMASHPSVSHAV 120
QY 121 GVSVMKASBAGAEYDDQTSOREKDDKVPFGSSHTYVQVLKENGPMASDPLCLTYSLSH 180
DB 121 GVSVMKASBAGAEYDDQTSOREKDDKVPFGSSHTYVQVLKENGPMASDPLCLTYSLSH 180
QY 181 VDLVKDNLNSGLIGALLVCREGSLAKEKTQTLHKFILLFAVPDEGKSWHSETKNSLMQORD 240
DB 181 VDLVKDNLNSGLIGALLVCREGSLAKEKTQTLHKFILLFAVPDEGKSWHSETKNSLMQORD 240
QY 241 AASARAPKMTVNGYVNRSLPGLIGCHRKSVYVHVIGMGTTPEVHSIFLEGHTFLVRNH 300
DB 241 AASARAPKMTVNGYVNRSLPGLIGCHRKSVYVHVIGMGTTPEVHSIFLEGHTFLVRNH 300
QY 301 ROASLEISPIITFLTAQTLMDLGOFLLSCHTSSSHOHGDMZAYVKVDSQPEPQIMKUNE 360
DB 301 ROASLEISPIITFLTAQTLMDLGOFLLSCHTSSSHOHGDMZAYVKVDSQPEPQIMKUNE 360
QY 361 EADYDDDLTDSEMDVVRPDDNSPSTQIRSVAKKHPTWVHYTAAEEDWDYAPLVLA 420
DB 361 EADYDDDLTDSEMDVVRPDDNSPSTQIRSVAKKHPTWVHYTAAEEDWDYAPLVLA 420
QY 421 PDRSYKSOYLNGPQIRGRYKVKRPNAYTDEFTKTEAACHESGILGLLYGEVGTLL 480
DB 421 PDRSYKSOYLNGPQIRGRYKVKRPNAYTDEFTKTEAACHESGILGLLYGEVGTLL 480
QY 481 LIIFKNOASRPVNIYPHGITDVRPLYSRRLPKGVKHLADFPILPGEIFKYKWTVTVEDCP 540
DB 481 LIIFKNOASRPVNIYPHGITDVRPLYSRRLPKGVKHLADFPILPGEIFKYKWTVTVEDCP 540
QY 541 TKSDPRCLTRYSSFWNNMERDLASGLIGPLICYKESVDQKQNMDSKRNVLFSVDE 600
DB 541 TKSDPRCLTRYSSFWNNMERDLASGLIGPLICYKESVDQKQNMDSKRNVLFSVDE 600
QY 601 NRSWYLTENIQRFENPAGVQLEDEPQASNMHSGINGVFPDSLOSLVCLHEVAYWYILS 660
DB 601 NRSWYLTENIQRFENPAGVQLEDEPQASNMHSGINGVFPDSLOSLVCLHEVAYWYILS 660
QY 661 IGAQTDFLSVFSGYTFXHKVYEDTLTFPFSGETVFMNMENPGLWILGCHNSDFRNRG 720
DB 661 IGAQTDFLSVFSGYTFXHKVYEDTLTFPFSGETVFMNMENPGLWILGCHNSDFRNRG 720
QY 721 MTALLKVSSCDKNTGDDYEDSDYEDISAYLLSKNNAIEPRSPQNSRHPSTROKQFNATTI 780
DB 721 MTALLKVSSCDKNTGDDYEDSDYEDISAYLLSKNNAIEPRSPQNSRHPSTROKQFNATTI 780
QY 781 PENDIEKTDPPFAHRTPMPKIQNVSSDILLMLROSPTPHGLSLDLOBAKYETETSDDES 840
DB 781 PENDIEKTDPPFAHRTPMPKIQNVSSDILLMLROSPTPHGLSLDLOBAKYETETSDDES 840
QY 841 PGALDSNNLSSEWTHFRPOLHESGDMVTFPESGLQRLNEKLGTTAATLKKLDFKVSST 900
DB 841 PGALDSNNLSSEWTHFRPOLHESGDMVTFPESGLQRLNEKLGTTAATLKKLDFKVSST 900
QY 901 SNNLSTIPSDNLAAGTDNTSSLGPPSMVHVDSQDITLFGKKSPLTESGGLSLSEE 960
DB 901 SNNLSTIPSDNLAAGTDNTSSLGPPSMVHVDSQDITLFGKKSPLTESGGLSLSEE 960
QY 961 NNDSKLLSGLMNSQESSWGKNVSTESGRLEPKGRAHGFALLTKDNALFKVSIULLKTN 1020
DB 961 NNDSKLLSGLMNSQESSWGKNVSTESGRLEPKGRAHGFALLTKDNALFKVSIULLKTN 1020
QY 1021 KTSNNSATNRKTHIDGPSLLIENSPSVWQNIILESSTEFKKVTPLIHDMMDKNATALEL 1080
DB 1021 KTSNNSATNRKTHIDGPSLLIENSPSVWQNIILESSTEFKKVTPLIHDMMDKNATALEL 1080
QY 1081 NEMSNKTTSSKKNMVOQKEGPIPPDAQNPDMSPFKMLFLPESARWIOETHCKSLNSG 1140

1081 NHMSNKTSSKNMVMQKKEGPPIPPDAQNDPMDSFFKWLFLIPESARVQRTGEGNSLNSG 1140
1141 QGSPKQVLSLGPBKSVEGQNFISEKNKVVGKEGFTKQVGLKEMVPPSSRNLPFLTNLDN 1200
1141 QGSPKQVLSLGPBKSVEGQNFISEKNKVVGKEGFTKQVGLKEMVPPSSRNLPFLTNLDN 1200
1201 LHENNTNHOQKKEOELEKKEKTELLIQENWVLPQIHVTGTGKNFMKNLFLLSRQNVESYD 1260
1201 LHENNTNHOQKKEOELEKKEKTELLIQENWVLPQIHVTGTGKNFMKNLFLLSRQNVESYD 1260
1261 GAYAPVLQDFRSLNDSNTRTKHTAHFSKKEEENLEGLGNQTKQIVKQVACTTRISPT 1320
1261 GAYAPVLQDFRSLNDSNTRTKHTAHFSKKEEENLEGLGNQTKQIVKQVACTTRISPT 1320
1321 SQQNFVQSRKALQKQRLPLEETLEKRIIIVDDTSQWNSNMKHLPESTLTQIDYKEKE 1380
1321 SQQNFVQSRKALQKQRLPLEETLEKRIIIVDDTSQWNSNMKHLPESTLTQIDYKEKE 1380
1381 KGAITQSPLSCLTRSHSIPOANRSPPIAKVSSFPSPRPIYLRVLPQDNSSHLPAASY 1440
1381 KGAITQSPLSCLTRSHSIPOANRSPPIAKVSSFPSPRPIYLRVLPQDNSSHLPAASY 1440
1441 RKDQGVQESHFTQGAQKNLNLALITLMTGQREVSGISGTSATNSVTVKKQVNTVLP 1500
1441 RKDQGVQESHFTQGAQKNLNLALITLMTGQREVSGISGTSATNSVTVKKQVNTVLP 1500
1501 KPD:PKTSGKVLLPKVHIYQKDLFPPTETSGSPCHLDLVEGSLQGTGEGAIKWEANRP 1560
1501 KPD:PKTSGKVLLPKVHIYQKDLFPPTETSGSPCHLDLVEGSLQGTGEGAIKWEANRP 1560
1561 GKVPFLRVATESSAKTSPKLDPLAWDNHYGTQIPKBEWKQESPEKTAFFKKDITLSL 1620
1561 GKVPFLRVATESSAKTSPKLDPLAWDNHYGTQIPKBEWKQESPEKTAFFKKDITLSL 1620
1621 NACSNHAIAINQGNKPEIEVTWAKQGRTERLCSONPPVLPKHQRHBITITLQSOQEE 1680
1621 NACSNHAIAINQGNKPEIEVTWAKQGRTERLCSONPPVLPKHQRHBITITLQSOQEE 1680
1681 IDYDDTTSVENKKEDFDYDDEDNQSPSPKTRHYEIRAEVRLWQYGMSSSHEVLNRP 1740
1681 IDYDDTTSVENKKEDFDYDDEDNQSPSPKTRHYEIRAEVRLWQYGMSSSHEVLNRP 1740
1741 AQSGSVPOFKKVVQFEPTDGSFTQPLRYGELNEHLGGLGPVIRAEVDNIMVTFRNQASR 1800
1741 AQSGSVPOFKKVVQFEPTDGSFTQPLRYGELNEHLGGLGPVIRAEVDNIMVTFRNQASR 1800
1801 PYSYSSLSIYEEBQORQAEPRKPKVKNETKTFWKVQHMAPTKDBDFCKAWAYPSDV 1860
1801 PYSYSSLSIYEEBQORQAEPRKPKVKNETKTFWKVQHMAPTKDBDFCKAWAYPSDV 1860
1861 DLEKDVHSLGIPLLVCHTNILNPAHQVQVTVQEPALFPTTFDETKSWYFTENNERCRA 1920
1861 DLEKDVHSLGIPLLVCHTNILNPAHQVQVTVQEPALFPTTFDETKSWYFTENNERCRA 1920
1921 PCNTQMEDPTTKENYRFAHNGYIMDTLPGLVMAQDQIRWYLISMGSGNENIHSIHFSGH 1980
1921 PCNTQMEDPTTKENYRFAHNGYIMDTLPGLVMAQDQIRWYLISMGSGNENIHSIHFSGH 1980
1981 VFTVRKKEEYKVALYNLYPGVPEIVEMLPKAGIWRVECLIGELHAGMSTLFLVYSNKC 2040
1981 VFTVRKKEEYKVALYNLYPGVPEIVEMLPKAGIWRVECLIGELHAGMSTLFLVYSNKC 2040
2041 QTPLGMAASHIRDFQITASQVQWAPKLARLHYSGSNANSTKBPFPFWIKVDLLAPMII 2100
2041 QTPLGMAASHIRDFQITASQVQWAPKLARLHYSGSNANSTKBPFPFWIKVDLLAPMII 2100
2101 HGKIQGARQKPFSSLYISQFIIMYSLDGKKWQTYRGNSTGLMWFFGNVDSGGIKHNIFN 2160
2101 HGKIQGARQKPFSSLYISQFIIMYSLDGKKWQTYRGNSTGLMWFFGNVDSGGIKHNIFN 2160
2161 PPIIARYIRLPHTHYSIRSLRMLMELMGDLNCSPLGMESKAIQITASSYFTNMFA 2220
2161 PPIIARYIRLPHTHYSIRSLRMLMELMGDLNCSPLGMESKAIQITASSYFTNMFA 2220

QY 2221 TWSPSKARLHLQGRSNARPOVANNPKEWLQVDFQTKMKVTGVTTOGVKSLTSMYVK3FL 2280
Db 2221 TWSPSKARLHLQGRSNARPOVANNPKEWLQVDFQTKMKVTGVTTOGVKSLTSMYVK3FL 2280
QY 2281 ISSSQDGHQWTLFFQNGKVKVFGQNGDSFTFPVNSLDPPILLTRYLRHPQSWVHQIALRM 2340
Db 2281 ISSSQDGHQWTLFFQNGKVKVFGQNGDSFTFPVNSLDPPILLTRYLRHPQSWVHQIALRM 2340
QY 2341 EVLGCEAQQDLY 2351
Db 2341 EVLGCEAQQDLY 2351

RESULT 4

US-08-121-202-2
; Sequence 2, Application US/09121202
; Patent No. 5563045
; GENERAL INFORMATION:
; APPLICANT: Pittman, Debra
; APPLICANT: Rehemtulla, Alnawaz
; APPLICANT: Wozney, John M.
; APPLICANT: Kaufman, Randal J.
; TITLE OF INVENTION: CHIMERIC PROCOAGULANT PROTEINS
; NUMBER OF SEQUENCES: 27
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Legal Affairs, Genetics Institute, Inc.
; STREET: 87 Cambridgepark Drive
; CITY: Cambridge
; STATE: MA
; COUNTRY: USA
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: JS/08/121.202
; FILING DATE: 14-SEP-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Meinert, M. C.
; REGISTRATION NUMBER: 31,544
; REFERENCE/DOCKET NUMBER: GI 5195A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)876-1210 X8574
; TELEFAX: (617) 876-5851
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2351 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-121-202-2

Query Match 99.8%; Score 12386; DB 1; Length 2351;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 2346; Conservative 2; Mismatches 3; Indels 0; Gaps 0;
QY 1 MQIELSTCFCLLRFCFSATRRYYLGAVELSWDMQSDLGELPVDARFPFPRVPKSPFFN 60
Db 1 MQIELSTCFCLLRFCFSATRRYYLGAVELSWDMQSDLGELPVDARFPFPRVPKSPFFN 60
QY 61 TSVVYKKTFLVEFTDHLFNIAKPPPMGLLGPITQAEVVDVTVTILKNKASHPVSLHAV 120
Db 61 TSVVYKKTFLVEFTDHLFNIAKPPPMGLLGPITQAEVVDVTVTILKNKASHPVSLHAV 120
QY 121 GVSWKASGEAYDDQTSQREKEDDKVFPGSGSHYTVVQVLKENGPMASDPLCLTYSYLSH 180
Db 121 GVSWKASGEAYDDQTSQREKEDDKVFPGSGSHYTVVQVLKENGPMASDPLCLTYSYLSH 180
QY 181 VDLVKDLNSGLIGALLVCREGSLAKEKTQTLHKFILLFAVDFEKGSHSETYNSLMDQDRD 240

Db 181 VDIYKOLNSGLIGALLVCREGLAKETQTLHKPILFAVDEGKSWHSETKNSIMQDRD 240
Qy 241 AASARAMPMTUNGYNRSIPGLIGCHRKVYVHVIGMGTTPVHSTFLFEGHTFLVRNH 300
Db 241 AASARAMPMTUNGYNRSIPGLIGCHRKVYVHVIGMGTTPVHSTFLFEGHTFLVRNH 300
Qy 301 ROASLEISPIFLTAQTLMLDGLFLLSCHLSSHODGMEAVKVDSPCEPQIMKONE 360
Db 301 ROASLEISPIFLTAQTLMLDGLFLLSCHLSSHODGMEAVKVDSPCEPQIMKONE 360
Qy 361 EAEYDDDLTDSMDVVRFDONSFSFIQIRSVAKKHPTWVHYIAABEEDWDYAPLVLA 420
Db 361 EAEYDDDLTDSMDVVRFDONSFSFIQIRSVAKKHPTWVHYIAABEEDWDYAPLVLA 420
Qy 421 PDRSRYKQYLNNQFORIGRKVKYRPMAYTDETEKTBREAIOHESGILGLLYGEGVDTL 480
Db 421 PDRSRYKQYLNNQFORIGRKVKYRPMAYTDETEKTBREAIOHESGILGLLYGEGVDTL 480
Qy 481 LIIIFKQASRPNIYPHGITDVRPLYSRRLPKGVKHLXDFPILPGEIYKYKWTVTVEDCP 540
Db 481 LIIIFKQASRPNIYPHGITDVRPLYSRRLPKGVKHLXDFPILPGEIYKYKWTVTVEDCP 540
Qy 541 TKSDPRCTRYIYSSFWNMRDLASGLIGLIIYKESYDQKQNMDSKRNWILFSVDE 600
Db 541 TKSDPRCTRYIYSSFWNMRDLASGLIGLIIYKESYDQKQNMDSKRNWILFSVDE 600
Qy 601 NRSWYLTENIQFLPNPAGVQLEDPEFOASNIHMSINGVFPDSLOLSVCLHEVAVWYILS 660
Db 601 NRSWYLTENIQFLPNPAGVQLEDPEFOASNIHMSINGVFPDSLOLSVCLHEVAVWYILS 660
Qy 661 IGAQDFLSVFPSSGYTFXHKWYVZTLTLPFSGTVMFMSMENPGLWILGCHNSDFNRG 720
Db 661 IGAQDFLSVFPSSGYTFXHKWYVZTLTLPFSGTVMFMSMENPGLWILGCHNSDFNRG 720
Qy 721 MTALLKVSQCDXNTGYEDSDYEDISAYLLSKNNAIEPRSPQNSKHSTROKQFNATTI 780
Db 721 MTALLKVSQCDXNTGYEDSDYEDISAYLLSKNNAIEPRSPQNSKHSTROKQFNATTI 780
Qy 781 PENDIEKTDPEFAHRTPMKIQNVSSDLMLLROSPTPHGLSLDLOEAKYETESDDPS 840
Db 781 PENDIEKTDPEFAHRTPMKIQNVSSDLMLLROSPTPHGLSLDLOEAKYETESDDPS 840
Qy 841 PGAISSNLSSEWTHFRPOLHESGDMVFTPEGLQRLNEXKLGTAAATELKKLDKFKVST 900
Db 841 PGAISSNLSSEWTHFRPOLHESGDMVFTPEGLQRLNEXKLGTAAATELKKLDKFKVST 900
Qy 901 SNNLITIPSDMLAAGTDNTSSLGPPSPVHYDSQDITLFGKKSPILTESGGPLSLSEE 960
Db 901 SNNLITIPSDMLAAGTDNTSSLGPPSPVHYDSQDITLFGKKSPILTESGGPLSLSEE 960
Qy 961 NNDKSLLESGLMNSQESSWGKNVSTESGRLFKGRHGPALLTXDNALFKVSIISLLKTN 1020
Db 961 NNDKSLLESGLMNSQESSWGKNVSTESGRLFKGRHGPALLTXDNALFKVSIISLLKTN 1020
Qy 1021 KTSNNSATNRKTHIDGPSLLIENSPSVWQNILESOTEFKKVTPLIHDMMLDKNATLRL 1080
Db 1021 KTSNNSATNRKTHIDGPSLLIENSPSVWQNILESOTEFKKVTPLIHDMMLDKNATLRL 1080
Qy 1081 NEMSNTTSSKNMWWQKKEGPIPPDAONPMSPEKMLFLPESAEWIORTHCKNSLNSG 1140
Db 1081 NEMSNTTSSKNMWWQKKEGPIPPDAONPMSPEKMLFLPESAEWIORTHCKNSLNSG 1140
Qy 1141 QGPPSPQVLGLSPEKVEGONFLSEKKNVVGKEFTKOVGLKEMVFPSSRNLFITNLND 1200
Db 1141 QGPPSPQVLGLSPEKVEGONFLSEKKNVVGKEFTKOVGLKEMVFPSSRNLFITNLND 1200
Qy 1201 LHEENTHNOEKTIQEBIEKKTILIQENVLPQIHTVTGTGNFMXNLLFLLSTRQNVGSYD 1260
Db 1201 LHEENTHNOEKTIQEBIEKKTILIQENVLPQIHTVTGTGNFMXNLLFLLSTRQNVGSYD 1260
Qy 1261 GAYAVILODFRSINDSTNRKTKHTAHSKKGSEENLEGIGNOTKQIVKCYACTRISPN 1320

Db 1261 GAYAVILODFRSINDSTNRKTKHTAHSKKGSEENLEGIGNOTKQIVKCYACTRISPN 1320
Qy 1321 SQQNFVTOREKRALQKQPLRP-EETELEKRIIVDDTSTQWSKNMKHLTPSTLTLOIDYNEKE 1380
Db 1321 SQQNFVTOREKRALQKQPLRP-EETELEKRIIVDDTSTQWSKNMKHLTPSTLTLOIDYNEKE 1380
Qy 1381 KGAITQSPLSDCTRSHSIPQANRSPPIAKVSSFPPIRPIYLRVLPQDNSSHLPAASY 1440
Db 1381 KGAITQSPLSDCTRSHSIPQANRSPPIAKVSSFPPIRPIYLRVLPQDNSSHLPAASY 1440
Qy 1441 RKDQSGVQESHSFLQGAKKONLSAILITLMTQDQREVSGLSATKSVYKVKVENTVLP 1500
Db 1441 RKDQSGVQESHSFLQGAKKONLSAILITLMTQDQREVSGLSATKSVYKVKVENTVLP 1500
Qy 1501 KPDLPKTSKGVKVELLPKHVHYKOLPPTETSNKSGPHGLDVEGSLLOQTEGAIKWNEANRP 1560
Db 1501 KPDLPKTSKGVKVELLPKHVHYKOLPPTETSNKSGPHGLDVEGSLLOQTEGAIKWNEANRP 1560
Qy 1561 GKVPFLRVATESSAKTFSKLLDPLAWNHYGTQPKBEWKSQESKPEKTAFFKKKDTILSL 1620
Db 1561 GKVPFLRVATESSAKTFSKLLDPLAWNHYGTQPKBEWKSQESKPEKTAFFKKKDTILSL 1620
Qy 1621 NACESNHAIAINBQONKPEIEVWAKQGRTERLCSQNPVLRKHOREITZTTTOSQOEE 1680
Db 1621 NACESNHAIAINBQONKPEIEVWAKQGRTERLCSQNPVLRKHOREITZTTTOSQOEE 1680
Qy 1681 IDYDDTISVENKKEDFDIYDEDNQSPRESPOKTRHYFIAAVERLWDYMGSSSPEVLRNR 1740
Db 1681 IDYDDTISVENKKEDFDIYDEDNQSPRESPOKTRHYFIAAVERLWDYMGSSSPEVLRNR 1740
Qy 1741 AQSGSVPOFKKVTFQEBFTDGSFTQLYRGEINELHGLLGPYIRAEVDNIMWTFRNQASR 1800
Db 1741 AQSGSVPOFKKVTFQEBFTDGSFTQLYRGEINELHGLLGPYIRAEVDNIMWTFRNQASR 1800
Qy 1801 PYSYSSLIISYEEBQRCGAERKMFVKNETKTYFWKVQOHHMPTKDFDCKAWAYPSDV 1860
Db 1801 PYSYSSLIISYEEBQRCGAERKMFVKNETKTYFWKVQOHHMPTKDFDCKAWAYPSDV 1860
Qy 1861 DLEKDVHSGLLGPLLVCHTNTLNPAHGRQVTVQBFALFTTIFDETKSWYFTENMERNCA 1920
Db 1861 DLEKDVHSGLLGPLLVCHTNTLNPAHGRQVTVQBFALFTTIFDETKSWYFTENMERNCA 1920
Qy 1921 PCNIQMEDPTPKENYRPHAINGYIMDTLPGVMAQDQIRWYLLSMGSENIHSHIFSGH 1980
Db 1921 PCNIQMEDPTPKENYRPHAINGYIMDTLPGVMAQDQIRWYLLSMGSENIHSHIFSGH 1980
Qy 1981 VFTVRKKEEYKXALYNLYPGYFETVEMLPKAGIWRVECLICEHLHAGWSTLFLVYSNKC 2040
Db 1981 VFTVRKKEEYKXALYNLYPGYFETVEMLPKAGIWRVECLICEHLHAGWSTLFLVYSNKC 2040
Qy 2041 QTPGLMASGHIRDFOITASGOYGQWAPKLARLHYSGSINAMSTKEPFSWIKVDLLAPMII 2100
Db 2041 QTPGLMASGHIRDFOITASGOYGQWAPKLARLHYSGSINAMSTKEPFSWIKVDLLAPMII 2100
Qy 2101 HGIKTCGARQKSFESSLYISQFIIMYSIDGKKWQYRGNSTGTLMVFFGNVDSSGKHNLN 2160
Db 2101 HGIKTCGARQKSFESSLYISQFIIMYSIDGKKWQYRGNSTGTLMVFFGNVDSSGKHNLN 2160
Qy 2161 PPIIARYIRLHPTHYISIRSTLRMELMGCDLNSCMPLGWESKAIISDAQITASSYFTNFA 2220
Db 2161 PPIIARYIRLHPTHYISIRSTLRMELMGCDLNSCMPLGWESKAIISDAQITASSYFTNFA 2220
Qy 2221 TMSPSKAEHLHQGSNARWPQNNPKWLQVDFOKTMKVTVTQGVKSLTSMYVKEFL 2280
Db 2221 TMSPSKAEHLHQGSNARWPQNNPKWLQVDFOKTMKVTVTQGVKSLTSMYVKEFL 2280
Qy 2281 TSSSQDGHQWTLFPQNGKVKVPOGQNSDFTPVVNSLDDPPLLTRYLRIHPOSWHQIALEM 2340
Db 2281 TSSSQDGHQWTLFPQNGKVKVPOGQNSDFTPVVNSLDDPPLLTRYLRIHPOSWHQIALEM 2340
Qy 2341 EVLIGCEAODLY 2351
Db 2341 EVLIGCEAODLY 2351

```
RESULT 5
5422260-1
;PATENT NO. 5422260
;APPLICANT: KAUFMAN, RANDAL J.;PITTMAN, DEBRA D.;TOOLE, JOHN J.
;TITLE OF INVENTION: HUMAN FACTOR VIII: C MUTAINS
;NUMBER OF SEQUENCES: 15
;CURRENT APPLICATION DATA:
;APPLICATION NUMBER: US/07/883,936
;FILING DATE: 15-MAY-1992
;PRIOR APPLICATION DATA:
;APPLICATION NUMBER: 279,485
;FILING DATE: 02-DEC-1988; 09-DEC-1986
;APPLICATION NUMBER: 939,658
;FILING DATE: 09-DEC-1986
;APPLICATION NUMBER: 932,767
;FILING DATE: 18-NOV-1986
;APPLICATION NUMBER: 868,410
;FILING DATE: 29-MAY-1986
;SEQ ID NO:1:
;LENGTH: 2351
5422260-1

Query Match 99.7%; Score 12377; DB 6; Length 2351;
Best Local Similarity 99.7%; Pred. No. 0;
Matches 2345; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

Qy 1 MQIELSTCFPLCLLRPCFSATRRYVLGAVELSDWMQSDLGELPVDARFPPRVKSPFPN 60
Db 1 MQIELSTCFPLCLLRPCFSATRRYVLGAVELSDWMQSDLGELPVDARFPPRVKSPFPN 60

Qy 61 TSVVYKKTFLVEFTDHLFNIAKPPNKGILGPLOAIVYTVITLKNASHVPVSLHAV 120
Db 61 TSVVYKKTFLVEFTVHLFNIAKPPNKGILGPLOAIVYTVITLKNASHVPVSLHAV 120

Qy 121 GVSVMKASEGAYDDOTSQRKEDDKVPPGSGHTYVWQVLKENGSMADPCLITYSYLSH 180
Db 121 GVSVMKASEGAYDDOTSQRKEDDKVPPGSGHTYVWQVLKENGSMADPCLITYSYLSH 180

Qy 181 VDLVKDLSGLIGALLVCREGSLAKEKQTQLHKFILLFAVDEGKSWHSETKNSLMQDRD 240
Db 181 VDLVKDLSGLIGALLVCREGSLAKEKQTQLHKFILLFAVDEGKSWHSETKNSLMQDRD 240

Qy 241 AASARAWPMHTVNGVYVNRSLPGLTGCHRSKYVYHVGWGTPEVHSTFLEHGTFLVRNH 300
Db 241 AASARAWPMHTVNGVYVNRSLPGLTGCHRSKYVYHVGWGTPEVHSTFLEHGTFLVRNH 300

Qy 301 RQASLEISPIITFLTAQTLIMDLGQFLLSCHISSHQHGDGMEAYKVVDSCPEBPQLMKNE 360
Db 301 RQASLEISPIITFLTAQTLIMDLGQFLLSCHISSHQHGDGMEAYKVVDSCPEBPQLMKNE 360

Qy 361 EAEDYDDDLTDSMDVVRPDDNSPFTQIRSVAKKPKTWVHYTAABEEDWDYAPLVLA 420
Db 361 EAEDYDDDLTDSMDVVRPDDNSPFTQIRSVAKKPKTWVHYTAABEEDWDYAPLVLA 420

Qy 421 PDRSVKSOYLNNGPQIRGRKVKYVRMAYTDETFKTRAIQHBGSLGLPALLYGVGDTL 480
Db 421 PDRSVKSOYLNNGPQIRGRKVKYVRMAYTDETFKTRAIQHBGSLGLPALLYGVGDTL 480

Qy 481 LLIIFKQASRPNIYPHGITDVRPYSRRLPKGVKELKDFPILPGEIKFYKWTVVEBGP 540
Db 481 LLIIFKQASRPNIYPHGITDVRPYSRRLPKGVKELKDFPILPGEIKFYKWTVVEBGP 540

Qy 541 TXSDPRCLTRYSSVFNBERDLASGLIGPLLICYKESVDQKQNMDSKRNWILSVFDE 600
Db 541 TXSDPRCLTRYSSVFNBERDLASGLIGPLLICYKESVDQKQNMDSKRNWILSVFDE 600

Qy 601 NESWYLTENIQRLNPAGVQLDEPFOASNMHESINGVVPDSLOLSVCLHVAWYIYLS 660
Db 601 NESWYLTENIQRLNPAGVQLDEPFOASNMHESINGVVPDSLOLSVCLHVAWYIYLS 660

Qy 661 IGAQTDFLSVFPGYTFKHVMYEDTLTLPFSGETVFMENPGLWILGCHNSDFRNRG 720
Db 661 IGAQTDFLSVFPGYTFKHVMYEDTLTLPFSGETVFMENPGLWILGCHNSDFRNRG 720

Qy 721 MTALLKVSSCDKNTGDIYEDSYEDISAYLLSKNNAIEPRSPQNSRHPSTROKQFNATTI 780
Db 721 MTALLKVSSCDKNTGDIYEDSYEDISAYLLSKNNAIEPRSPQNSRHPSTROKQFNATTI 780

Qy 781 PENDIEKTDPEAHRTPMPKIQNVSSDDLMLLRQSPTRHGLSLSDLOEAKYEFSDPS 840
Db 781 PENDIEKTDPEAHRTPMPKIQNVSSDDLMLLRQSPTRHGLSLSDLOEAKYEFSDPS 840

Qy 841 PGALDSNNLSSEMTFRPQLHSGDMVFTPESGIQLRLNEKLGTTAATELKKLDFKVSST 900
Db 841 PGALDSNNLSSEMTFRPQLHSGDMVFTPESGIQLRLNEKLGTTAATELKKLDFKVSST 900

Qy 901 SNNLSTIPSDNLAAGTDNTSSILGPPMPVHYDSOLDTTLFGKSSPLTSGGFLSSEE 960
Db 901 SNNLSTIPSDNLAAGTDNTSSILGPPMPVHYDSOLDTTLFGKSSPLTSGGFLSSEE 960

Qy 961 NNDKLLLESGIMNSQESSWGKNVSTESGRLEFKGRAHGPAALLTKDNALFKVSTLSLKTN 1020
Db 961 NNDKLLLESGIMNSQESSWGKNVSTESGRLEFKGRAHGPAALLTKDNALFKVSTLSLKTN 1020

Qy 1021 KTSNNSATNRKTHIDGFSLLIENSPPVWQNLLESDETFKKVTPLIHDMMLMDKNATALRL 1080
Db 1021 KTSNNSATNRKTHIDGFSLLIENSPPVWQNLLESDETFKKVTPLIHDMMLMDKNATALRL 1080

Qy 1081 NHMKNKTISSKNMHWQKKEGPIPPDAQNPDSFFKMLFLPESARWQRTGKNSLNSG 1140
Db 1081 NHMKNKTISSKNMHWQKKEGPIPPDAQNPDSFFKMLFLPESARWQRTGKNSLNSG 1140

Qy 1141 QGPPKQNLISLGPESVEGQNFSEKNKVVVGKEFTKDVGLKEMVFPSSRNFLTNLDN 1200
Db 1141 QGPPKQNLISLGPESVEGQNFSEKNKVVVGKEFTKDVGLKEMVFPSSRNFLTNLDN 1200

Qy 1201 LHENNTNHOEKKIQEEIEKKEKTELIOENVLPQIHVTGTGNFMKNLFLSTRQNVESYE 1260
Db 1201 LHENNTNHOEKKIQEEIEKKEKTELIOENVLPQIHVTGTGNFMKNLFLSTRQNVESYE 1260

Qy 1261 GAYAPVLQDPRSLNDSINRHKHTAHFSKKEENLEGLNQTKQIVEKYACTTRISNT 1320
Db 1261 GAYAPVLQDPRSLNDSINRHKHTAHFSKKEENLEGLNQTKQIVEKYACTTRISNT 1320

Qy 1321 SQQNFVQSRKALKQRLPLEETELEKRIIVDDTSTQWSNMKHLTPSTLTQIDYNEKE 1380
Db 1321 SQQNFVQSRKALKQRLPLEETELEKRIIVDDTSTQWSNMKHLTPSTLTQIDYNEKE 1380

Qy 1381 KGATQSPSLDCLTRSHSIPQANRSPPIAKVSSPPSIRPIYLTRVLPODNSSHLPAASY 1440
Db 1381 KGATQSPSLDCLTRSHSIPQANRSPPIAKVSSPPSIRPIYLTRVLPODNSSHLPAASY 1440

Qy 1441 RKDQSGVQESSHFLQGAKNKNSLAILTLEMTGQREVSGISGTSATNSVTYKVENTVLP 1500
Db 1441 RKDQSGVQESSHFLQGAKNKNSLAILTLEMTGQREVSGISGTSATNSVTYKVENTVLP 1500

Qy 1501 KPDLPKTSKGVKELLPKVHIYQKDLFPPTETSNSSFGHLDLVEGSLQGTGEGAIKWEANRP 1560
Db 1501 KPDLPKTSKGVKELLPKVHIYQKDLFPPTETSNSSFGHLDLVEGSLQGTGEGAIKWEANRP 1560

Qy 1561 GKVPFLRVATSSAKTPSKLLDPLAWNHGYTOIPKEWKQKESPEKTAFAKKKDTILSL 1620
Db 1561 GKVPFLRVATSSAKTPSKLLDPLAWNHGYTOIPKEWKQKESPEKTAFAKKKDTILSL 1620

Qy 1621 NACSNHAIAINSGONKPEITEVWAKQGRTERLCSQNPPLVKRHOEITRTTILQSDOE 1680
Db 1621 NACSNHAIAINSGONKPEITEVWAKQGRTERLCSQNPPLVKRHOEITRTTILQSDOE 1680

Qy 1681 IDYDDTIQSVEMKKEDFDIYDEDNQSPRSFQKTRHYPIAAVERLWDYCMSSSPHLNRR 1740
Db 1681 IDYDDTIQSVEMKKEDFDIYDEDNQSPRSFQKTRHYPIAAVERLWDYCMSSSPHLNRR 1740

Qy 1741 AQSQSVDPQFKVYQEFQFTDGSFTQPLYEGELNEHLGLLGPYIRAVEEDINIMVTFNQASR 1800
Db 1741 AQSQSVDPQFKVYQEFQFTDGSFTQPLYEGELNEHLGLLGPYIRAVEEDINIMVTFNQASR 1800
```

1741 AQSQVDPQKVVQFEPFTDGSFTQPLRGELNEHLGLGPVIRAEVDNIMVIFRNQASR 1800
1801 PYSYSSLSIYEEBQORQAERKPKFVKNEFKTYFWKQCHMAPTKDEFDCANAYPSDV 1860
1801 PYSYSSLSIYEEBQORQAERKPKFVKNEFKTYFWKQCHMAPTKDEFDCANAYPSDV 1860
1861 DLEKDVHSLGLPLVCHTNTLNFAHQVQVVFQFALPFTTFDETKSWYFTENNERCRA 1920
1861 DLEKDVHSLGLPLVCHTNTLNFAHQVQVVFQFALPFTTFDETKSWYFTENNERCRA 1920
1921 PCNIQMEDPTKFNRYFHAINGYIMDTLPLGLVMAQDQIRWYLLSMNGENIHSIHFSGH 1980
1921 PCN-QMEDPTKFNRYFHAINGYIMDTLPLGLVMAQDQIRWYLLSMNGENIHSIHFSGH 1980
1981 VFTVRKKEEYKVALNLYPGVFPETVEMLPKAGIWRVECLIGELHAGWSTLFLVYSNKC 2040
1981 VFTVRKKEEYKVALNLYPGVFPETVEMLPKAGIWRVECLIGELHAGWSTLFLVYSNKC 2040
2041 QTPLGWASGHIRDFOITASGOYGQWAPKRLARLHYSGSINAWSTKEPFSWIKVDLIAPMII 2100
2041 QTPLGWASGHIRDFOITASGOYGQWAPKRLARLHYSGSINAWSTKEPFSWIKVDLIAPMII 2100
2101 HGKITQARQKPSLSYISQFIIMYSLDCKKQTYRGNSTGILMVPFGNVDSGKHNIFN 2160
2101 HGKITQARQKPSLSYISQFIIMYSLDCKKQTYRGNSTGILMVPFGNVDSGKHNIFN 2160
2161 PPIIARYRLPHYTHYSIRSLRMELMGCDLNSCPLGNESKAISDAQITASSYFTNMFA 2220
2161 PPIIARYRLPHYTHYSIRSLRMELMGCDLNSCPLGNESKAISDAQITASSYFTNMFA 2220
2221 TWSPSKARLHJQGRSNARWPNVNNKWLQVDFQKTMKVTTQGVKSLTSMYVKEFL 2280
2221 TWSPSKARLHJQGRSNARWPNVNNKWLQVDFQKTMKVTTQGVKSLTSMYVKEFL 2280
2281 ISSODQCHQWTLFFQNGKVKVQFQGNQDSFTPVNSLDPLTRYLRIRHPOSVHQIALRM 2340
2281 ISSODQCHQWTLFFQNGKVKVQFQGNQDSFTPVNSLDPLTRYLRIRHPOSVHQIALRM 2340
2341 EVLGCEAQDLY 2351
2341 EVLGCEAQDLY 2351

RESULT 6

US-07-864-004B-4
Sequence 4, Application US/07864004B
Patent No. 5364771
GENERAL INFORMATION:
APPLICANT: Lollar, John S.
APPLICANT: Runge, Marshall S.
TITLE OF INVENTION: Hybrid Human/Porcine Factor VIII
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSER: Kilpatrick & Cody
STREET: 1100 Peachtree Street
CITY: Atlanta
STATE: Georgia
COUNTRY: US
ZIP: 30309
COMPUTER READABLE FORM: disk
MEDIUM TYPE: Floppy
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/864,004B
FILING DATE: 07 APRIL 1992
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Pabst, Patrea L.
REGISTRATION NUMBER: 31,284
REFERENCE/DOCKET NUMBER: EMU106
TELECOMMUNICATION INFORMATION:

TELEPHONE: 404-815-6508
TELEFAX: 404-815-6855
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 2332 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE: N-terminal
ORIGINAL SOURCE:
ORGANISM: Homo sapien
TISSUE TYPE: Liver cDNA sequence
US-07-864-004B-4
Query Match 98.9%; Score 12282; DB 1; Length 2332;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 2327; Conservative 2; Mismatches 3; Indels 0; Gaps 0;
QY 20 ATRVYVIGAVELSDYMSDGLGELPVDARPPRPVKSPPTNTSVVYKTLFVEFTDHLFN 79
DB 1 ATRVYVIGAVELSDYMSDGLGELPVDARPPRPVKSPPTNTSVVYKTLFVEFTDHLFN 60
QY 80 IAKPRPPMGLLPTIOAEVYDVTVITLKNVASHPVSLHAGVSVYWKASGEAYDDQTSQ 139
DB 61 IAKPRPPMGLLPTIOAEVYDVTVITLKNVASHPVSLHAGVSVYWKASGEAYDDQTSQ 120
QY 140 REKEDDKVFPGGSTTYWQVLKENGPMASDPLCTYSVLSVDLVKDLNSGLIGALLVCR 199
DB 121 REKEDDKVFPGGSTTYWQVLKENGPMASDPLCTYSVLSVDLVKDLNSGLIGALLVCR 180
QY 200 EGS LAKEKTOFLHPIILLFAVDFDGKSWHSTKNSLMODRDAAGARAWPKVHTVNGYVNR 259
DB 181 EGS LAKEKTOFLHPIILLFAVDFDGKSWHSTKNSLMODRDAAGARAWPKVHTVNGYVNR 240
QY 260 SLPLGLIGCHRSKVYVHVGCTTPEVHSIFLEGHTFLVRNHRQASLEISP-TFLTAQTLL 319
DB 241 SLPLGLIGCHRSKVYVHVGCTTPEVHSIFLEGHTFLVRNHRQASLEISP-TFLTAQTLL 300
QY 320 MDLQGPLLSCHISSHQHDGMEAYVKVDCSPEPQILMKNNBEADYDDDLTDSMDVYRF 379
DB 301 MDLQGPLLSCHISSHQHDGMEAYVKVDCSPEPQILMKNNBEADYDDDLTDSMDVYRF 360
QY 380 DDNSPSPFIQIRSVAKKPKTWVHYIAAEEDWDYAPLVAFAPDDRYSYKSOYLNNGPQIG 439
DB 361 DDNSPSPFIQIRSVAKKPKTWVHYIAAEEDWDYAPLVAFAPDDRYSYKSOYLNNGPQIG 420
QY 440 RYKVKVRPMAYTDETFKTRTREAIOHESGILGPLLYGVEGDTLLIIPKQASRPYNIYPHGI 499
DB 421 RYKVKVRPMAYTDETFKTRTREAIOHESGILGPLLYGVEGDTLLIIPKQASRPYNIYPHGI 480
QY 500 TDVREPLYSRRLPKGVKHLKDFPILPGEIFKYKWTVTVEDGPKSDPRCLTYYSFVAME 559
DB 481 TDVREPLYSRRLPKGVKHLKDFPILPGEIFKYKWTVTVEDGPKSDPRCLTYYSFVAME 540
QY 560 RDLASGLIGPLLIICYKESVDOKNGQIMGDKRNVILFSVFDENRSWYLTENIQRFLENPAG 619
DB 541 RDLASGLIGPLLIICYKESVDOKNGQIMGDKRNVILFSVFDENRSWYLTENIQRFLENPAG 600
QY 620 VOLEDPFQASNIHNSINGYVFDLSIQLSVCLHEVAYWYILSIGAOTDPLSFVFFSGYTFKH 679
DB 601 VOLEDPFQASNIHNSINGYVFDLSIQLSVCLHEVAYWYILSIGAOTDPLSFVFFSGYTFKH 660
QY 680 KMVYEDTLTLFPFSGETVMSXENPGLWILGCHNSDPENRGMTALLKYSSCDKNTGDYYE 739
DB 661 KMVYEDTLTLFPFSGETVMSXENPGLWILGCHNSDPENRGMTALLKYSSCDKNTGDYYE 720
QY 740 DSYEDISAYLLSKNNAIEPRSPSQNSRHFSTQKQFNATTPENDIEKTDPAFAHRTMP 799
DB 721 DSYEDISAYLLSKNNAIEPRSPSQNSRHFSTQKQFNATTPENDIEKTDPAFAHRTMP 780

800 KIQNVSSDILLMLLRQSPTRHGLSLSDQBAKYETTFSDPSPGALDSNNSLSEMTHERPQ 859
801 KIQNVSSDILLMLLRQSPTRHGLSLSDQBAKYETTFSDPSPGALDSNNSLSEMTHERPQ 840
802 KIQNVSSDILLMLLRQSPTRHGLSLSDQBAKYETTFSDPSPGALDSNNSLSEMTHERPQ 840
803 LHHGDMVFTPEGLQJRLNEKLGTAATELKKLDFKVSSTSNLIPTSDNLAAGTDN 919
804 LHHGDMVFTPEGLQJRLNEKLGTAATELKKLDFKVSSTSNLIPTSDNLAAGTDN 900
805 TSSLGSPMPHYDSOLDTTLFGKXSSPLTBSGGPLSLSEBNDKSKLLESGLNNSQESSW 979
806 TSSLGSPMPHYDSOLDTTLFGKXSSPLTBSGGPLSLSEBNDKSKLLESGLNNSQESSW 960
807 TSSLGSPMPHYDSOLDTTLFGKXSSPLTBSGGPLSLSEBNDKSKLLESGLNNSQESSW 960
808 GKNVSSPEGLFKGKRAHGALLATKCNALFKVSIKTKNKTSNNSATNKTTHIDGSL 1039
809 GKNVSSPEGLFKGKRAHGALLATKCNALFKVSIKTKNKTSNNSATNKTTHIDGSL 1020
810 GKNVSSPEGLFKGKRAHGALLATKCNALFKVSIKTKNKTSNNSATNKTTHIDGSL 1020
811 LIENSPVQWNLLESDETFKVTPLIHRMLMDKNATRLRNHNSKTTSSKNMEMVQOK 1099
812 LIENSPVQWNLLESDETFKVTPLIHRMLMDKNATRLRNHNSKTTSSKNMEMVQOK 1080
813 KEGPIPPDAQNPDMSEFFKVLFLPESARWIQTHGKNSLNSGQSPKQVLVSLGPEKSVEG 1159
814 KEGPIPPDAQNPDMSEFFKVLFLPESARWIQTHGKNSLNSGQSPKQVLVSLGPEKSVEG 1140
815 KEGPIPPDAQNPDMSEFFKVLFLPESARWIQTHGKNSLNSGQSPKQVLVSLGPEKSVEG 1140
816 QNFLSEKKNVVGGEFTKDVGLKEMVPPSSRNFLTLNOLHNNHNTNHOEKKIQEIEK 1219
817 QNFLSEKKNVVGGEFTKDVGLKEMVPPSSRNFLTLNOLHNNHNTNHOEKKIQEIEK 1200
818 KETLIOENNVLPQHTVTGTGNFKMKNLFLSTRQNVGSGYAGAPVLQDPRSLNDSTNR 1279
819 KETLIOENNVLPQHTVTGTGNFKMKNLFLSTRQNVGSGYAGAPVLQDPRSLNDSTNR 1260
820 TKKHTAFHSGKEBENLEGLNQTKQIVEXYACVATRIISPNTSQNFVQSRKALKQFRL 1339
821 TKKHTAFHSGKEBENLEGLNQTKQIVEXYACVATRIISPNTSQNFVQSRKALKQFRL 1320
822 PLEETELEKRIIVDDTSTQWKNMKHLPSTLTQIDYNEKEKGAITQSPSLSDCLTRSHSI 1399
823 PLEETELEKRIIVDDTSTQWKNMKHLPSTLTQIDYNEKEKGAITQSPSLSDCLTRSHSI 1380
824 POANRSLPIAKVSSPFIPIYITRVLPONSHLPAASTRKDSGVQSSHFLQGAKK 1459
825 POANRSLPIAKVSSPFIPIYITRVLPONSHLPAASTRKDSGVQSSHFLQGAKK 1440
826 NNLSLAILTEMTGQREVGSLGTSATNSVYKVENVLPKPLPKTSGXVLLPKVHI 1519
827 NNLSLAILTEMTGQREVGSLGTSATNSVYKVENVLPKPLPKTSGXVLLPKVHI 1500
828 YQKOLFPETENGSGHLLDVEGSLLOQTEGAIKWNEANRPGKVPFLRVATESAKTPSK 1579
829 YQKOLFPETENGSGHLLDVEGSLLOQTEGAIKWNEANRPGKVPFLRVATESAKTPSK 1560
830 LLDPLANDNHYGTQIPKEWKSQSKSPKTAFFKDDTILSNACESHAIINNEGQNK 1639
831 LLDPLANDNHYGTQIPKEWKSQSKSPKTAFFKDDTILSNACESHAIINNEGQNK 1620
832 EIEVTWAKQRTBLCSNPPVLKXHQSEIARTTLQSDQESIDYDDTISVPMKXEDFDIY 1699
833 EIEVTWAKQRTBLCSNPPVLKXHQSEIARTTLQSDQESIDYDDTISVPMKXEDFDIY 1680
834 DEDNQSPRSQKTRHYFIAAVERLWDYGMSSSPHVLNRNAQSGSVQPKVQVQFTD 1759
835 DEDNQSPRSQKTRHYFIAAVERLWDYGMSSSPHVLNRNAQSGSVQPKVQVQFTD 1740
836 GSFTQPLVGEINHLGLGPIYIAEVEDNTMTFRNQASPYFYSSLISYEDQROGA 1819
837 GSFTQPLVGEINHLGLGPIYIAEVEDNTMTFRNQASPYFYSSLISYEDQROGA 1800
838 EPRKNFVKNETKTYFWKVQHHMPTKDEPCKAWAYPSDVLKDVHSGHLIGLILVCHT 1879
839 EPRKNFVKNETKTYFWKVQHHMPTKDEPCKAWAYPSDVLKDVHSGHLIGLILVCHT 1860
840 NTLNPAHGRQVTVQEFALFTTIFDETSWFTENMERNCRAPCNIQMEDPTFKENYRPHA 1939

800 KIQNVSSDILLMLLRQSPTRHGLSLSDQBAKYETTFSDPSPGALDSNNSLSEMTHERPQ 859
801 KIQNVSSDILLMLLRQSPTRHGLSLSDQBAKYETTFSDPSPGALDSNNSLSEMTHERPQ 840
802 KIQNVSSDILLMLLRQSPTRHGLSLSDQBAKYETTFSDPSPGALDSNNSLSEMTHERPQ 840
803 LHHGDMVFTPEGLQJRLNEKLGTAATELKKLDFKVSSTSNLIPTSDNLAAGTDN 919
804 LHHGDMVFTPEGLQJRLNEKLGTAATELKKLDFKVSSTSNLIPTSDNLAAGTDN 900
805 TSSLGSPMPHYDSOLDTTLFGKXSSPLTBSGGPLSLSEBNDKSKLLESGLNNSQESSW 979
806 TSSLGSPMPHYDSOLDTTLFGKXSSPLTBSGGPLSLSEBNDKSKLLESGLNNSQESSW 960
807 TSSLGSPMPHYDSOLDTTLFGKXSSPLTBSGGPLSLSEBNDKSKLLESGLNNSQESSW 960
808 GKNVSSPEGLFKGKRAHGALLATKCNALFKVSIKTKNKTSNNSATNKTTHIDGSL 1039
809 GKNVSSPEGLFKGKRAHGALLATKCNALFKVSIKTKNKTSNNSATNKTTHIDGSL 1020
810 GKNVSSPEGLFKGKRAHGALLATKCNALFKVSIKTKNKTSNNSATNKTTHIDGSL 1020
811 LIENSPVQWNLLESDETFKVTPLIHRMLMDKNATRLRNHNSKTTSSKNMEMVQOK 1099
812 LIENSPVQWNLLESDETFKVTPLIHRMLMDKNATRLRNHNSKTTSSKNMEMVQOK 1080
813 KEGPIPPDAQNPDMSEFFKVLFLPESARWIQTHGKNSLNSGQSPKQVLVSLGPEKSVEG 1159
814 KEGPIPPDAQNPDMSEFFKVLFLPESARWIQTHGKNSLNSGQSPKQVLVSLGPEKSVEG 1140
815 KEGPIPPDAQNPDMSEFFKVLFLPESARWIQTHGKNSLNSGQSPKQVLVSLGPEKSVEG 1140
816 QNFLSEKKNVVGGEFTKDVGLKEMVPPSSRNFLTLNOLHNNHNTNHOEKKIQEIEK 1219
817 QNFLSEKKNVVGGEFTKDVGLKEMVPPSSRNFLTLNOLHNNHNTNHOEKKIQEIEK 1200
818 KETLIOENNVLPQHTVTGTGNFKMKNLFLSTRQNVGSGYAGAPVLQDPRSLNDSTNR 1279
819 KETLIOENNVLPQHTVTGTGNFKMKNLFLSTRQNVGSGYAGAPVLQDPRSLNDSTNR 1260
820 TKKHTAFHSGKEBENLEGLNQTKQIVEXYACVATRIISPNTSQNFVQSRKALKQFRL 1339
821 TKKHTAFHSGKEBENLEGLNQTKQIVEXYACVATRIISPNTSQNFVQSRKALKQFRL 1320
822 PLEETELEKRIIVDDTSTQWKNMKHLPSTLTQIDYNEKEKGAITQSPSLSDCLTRSHSI 1399
823 PLEETELEKRIIVDDTSTQWKNMKHLPSTLTQIDYNEKEKGAITQSPSLSDCLTRSHSI 1380
824 POANRSLPIAKVSSPFIPIYITRVLPONSHLPAASTRKDSGVQSSHFLQGAKK 1459
825 POANRSLPIAKVSSPFIPIYITRVLPONSHLPAASTRKDSGVQSSHFLQGAKK 1440
826 NNLSLAILTEMTGQREVGSLGTSATNSVYKVENVLPKPLPKTSGXVLLPKVHI 1519
827 NNLSLAILTEMTGQREVGSLGTSATNSVYKVENVLPKPLPKTSGXVLLPKVHI 1500
828 YQKOLFPETENGSGHLLDVEGSLLOQTEGAIKWNEANRPGKVPFLRVATESAKTPSK 1579
829 YQKOLFPETENGSGHLLDVEGSLLOQTEGAIKWNEANRPGKVPFLRVATESAKTPSK 1560
830 LLDPLANDNHYGTQIPKEWKSQSKSPKTAFFKDDTILSNACESHAIINNEGQNK 1639
831 LLDPLANDNHYGTQIPKEWKSQSKSPKTAFFKDDTILSNACESHAIINNEGQNK 1620
832 EIEVTWAKQRTBLCSNPPVLKXHQSEIARTTLQSDQESIDYDDTISVPMKXEDFDIY 1699
833 EIEVTWAKQRTBLCSNPPVLKXHQSEIARTTLQSDQESIDYDDTISVPMKXEDFDIY 1680
834 DEDNQSPRSQKTRHYFIAAVERLWDYGMSSSPHVLNRNAQSGSVQPKVQVQFTD 1759
835 DEDNQSPRSQKTRHYFIAAVERLWDYGMSSSPHVLNRNAQSGSVQPKVQVQFTD 1740
836 GSFTQPLVGEINHLGLGPIYIAEVEDNTMTFRNQASPYFYSSLISYEDQROGA 1819
837 GSFTQPLVGEINHLGLGPIYIAEVEDNTMTFRNQASPYFYSSLISYEDQROGA 1800
838 EPRKNFVKNETKTYFWKVQHHMPTKDEPCKAWAYPSDVLKDVHSGHLIGLILVCHT 1879
839 EPRKNFVKNETKTYFWKVQHHMPTKDEPCKAWAYPSDVLKDVHSGHLIGLILVCHT 1860
840 NTLNPAHGRQVTVQEFALFTTIFDETSWFTENMERNCRAPCNIQMEDPTFKENYRPHA 1939

RESULT 7

US-08-251-937A-4
; Sequence 4, Application: US/08251937A
; Patent No. 5583209
GENERAL INFORMATION:
APPLICANT: Lollar, John S.
APPLICANT: Rudge, Marshall S.
TITLE OF INVENTION: Hybrid Human/Porcine Factor VIII
NUMBER OF SEQUENCES: 13
CORRESPONDENCE ADDRESS:
ADDRESSEE: Kilpatrick & Cody
STREET: 1100 Peachtree Street
CITY: Atlanta
STATE: Georgia
COUNTRY: US
ZIP: 30309
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatenIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/251,937A
FILING DATE: 31-MAY-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/864,004
FILING DATE: 07-APR-1992
ATTORNEY/AGENT INFORMATION:
NAME: Pratt, John S.
REGISTRATION NUMBER: 29,476
REFERENCE/DOCKET NUMBER: EMU106DIV
TELECOMMUNICATION INFORMATION:
TELEPHONE: 404-815-6367
TELEFAX: 404-815-6555
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 2332 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear

																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																						</	
--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	----	--

Db 1981 GVFETVEMLPKXAGINRVECLHSHLAGMSTFLVYVSNKQCPPLGWASGHIRDFOITAS 2040
QY 2060 GOYGOBAPKILARLHYSGSINAWTKKPPFWIKVDLLAPMIIHGKIKTOGAROKFSSLYISQ 2119
Db 2041 GOYGOBAPKILARLHYSGSINAWTKKPPFWIKVDLLAPMIIHGKIKTOGAROKFSSLYISQ 2100
QY 2120 FIIMYSLDGKKWOTYRGNSTGTLWVFGNVDSSGIGKINENPPIIARYIRLHPHYSIRS 2179
Db 2101 FIIMYSLDGKKWOTYRGNSTGTLWVFGNVDSSGIGKINENPPIIARYIRLHPHYSIRS 2160
QY 2180 TLMELMGCDLNSCMLPLGMSKAISDAQITASSYFTNMFWATWSPSKARLHLCGRSNMWR 2239
Db 2161 TLMELMGCDLNSCMLPLGMSKAISDAQITASSYFTNMFWATWSPSKARLHLCGRSNMWR 2220
QY 2240 PQVNNPKWLQVDFQKTMKVGTGTVQGVKSLLSVMYVKEFLISSODGCHQWTLFFQNGKV 2299
Db 2221 PQVNNPKWLQVDFQKTMKVGTGTVQGVKSLLSVMYVKEFLISSODGCHQWTLFFQNGKV 2280
QY 2300 KVFQGNQDSFTPVVNSLDPLLTRYLRIHPQSWHQAIAEVEVLGCEAODLY 2351
Db 2281 KVFQGNQDSFTPVVNSLDPLLTRYLRIHPQSWHQAIAEVEVLGCEAODLY 2332

RESULT 8

US-08-212-133A-2
; Sequence 2, Application US/08212133A
; Patent No. 5663060
; GENERAL INFORMATION:
; APPLICANT: Lollar, John S.
; APPLICANT: Runge, Marschall S.
; TITLE OF INVENTION: Hybrid Human/Animal Factor VIII
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Kilpatrick & Cody
; STREET: 100 Peachtree Street
; CITY: Atlanta
; STATE: Georgia
; COUNTRY: US
; ZIP: 30303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/212,133A
; FILING DATE: March 11, 1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/864,004
; FILING DATE: 07-APR-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Pabst, Patrea L.
; REGISTRATION NUMBER: 31,284
; REFERENCE/DOCKET NUMBER: EMU/76677
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 404-572-6508
; TELEFAX: 404-572-6555
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2332 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHEetical: YES
; ANTI-SENSE: NO
; FRAGMENT TYPE: N-terminal
; ORIGINAL SOURCE:
; ORGANISM: Homo sapien
; TISSUE TYPE: Liver CDNA sequence
US-08-212-133A-2

Query Match 98.9%; Score 12282; DB 1; Length 2332;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 2327; Conservative 2; Mismatches 3; Indels 0; Gaps 0;
QY 20 ATRRYVLGAVELSWDVMQSDLGELPVDARPPRVPKSPFFNTSVVYKTLFVFTDHLFN 79
Db 1 ATRRYVLGAVELSWDVMQSDLGELPVDARPPRVPKSPFFNTSVVYKTLFVFTDHLFN 60
QY 80 IAKPRPPWMLLGPPTQAEVYDTWITLKNMASHPVSLHVGVSWKASGAEYDDOTSQ 139
Db 61 IAKPRPPWMLLGPPTQAEVYDTWITLKNMASHPVSLHVGVSWKASGAEYDDOTSQ 120
QY 140 REXEDDKVPFGGSHTYVWVVLKENGPMASDPLCLTYSYLSHVDLVKDLNSGLLGALLVCR 199
Db 121 REXEDDKVPFGGSHTYVWVVLKENGPMASDPLCLTYSYLSHVDLVKDLNSGLLGALLVCR 180
QY 200 EGS LAKEKTQTLHKFILLFAVDFEGKSWHSETKNSLMQDRDAASARAWPKQHTVNGVNR 259
Db 181 EGS LAKEKTQTLHKFILLFAVDFEGKSWHSETKNSLMQDRDAASARAWPKQHTVNGVNR 240
QY 260 SIPGLIGCHRKSVYWHVIGNGITPEVHSIFLEGHTFLVRNHRQASLEISPIITLTAQTLL 319
Db 241 SLPLIGCHRKSVYWHVIGNGITPEVHSIFLEGHTFLVRNHRQASLEISPIITLTAQTLL 300
QY 320 MDLGQFLSCHISSHQHDGMEAYVKYVDSCEPEEPOLIMKNEEAEYDDDLTDSMDVVRP 379
Db 301 MDLGQFLSCHISSHQHDGMEAYVKYVDSCEPEEPOLIMKNEEAEYDDDLTDSMDVVRP 360
QY 380 DDNSPSFIQIRSVAKKHPKTVWHYIAAEEDWDYAPLVLPDDRYSKYSLNNGPQIRIG 439
Db 361 DDNSPSFIQIRSVAKKHPKTVWHYIAAEEDWDYAPLVLPDDRYSKYSLNNGPQIRIG 420
QY 440 RYKVKRFWAYTDETPKTBRAIOHESGILGLPLYGEGVDTLLIIFKNOASRPNIYPHGI 499
Db 421 RYKVKRFWAYTDETPKTBRAIOHESGILGLPLYGEGVDTLLIIFKNOASRPNIYPHGI 480
QY 500 TDVRPLYSRRLPKGVKHLKDFPILPGEIFKYKVTVTVEGFTKSDPRCLTRYSSFVME 559
Db 481 TDVRPLYSRRLPKGVKHLKDFPILPGEIFKYKVTVTVEGFTKSDPRCLTRYSSFVME 540
QY 560 RDLASGLIGPLLCYKESVDQKGNQIMSKRNVLFSVFDENRSWYITENIQIFLENPAG 619
Db 541 RDLASGLIGPLLCYKESVDQKGNQIMSKRNVLFSVFDENRSWYITENIQIFLENPAG 600
QY 620 VOLDEPFOASNTMHSINGVVFDSQLSVCLHVAWYIILSICAOQDFLSVFFSGTIFKH 679
Db 601 VOLDEPFOASNTMHSINGVVFDSQLSVCLHVAWYIILSICAOQDFLSVFFSGTIFKH 660
QY 680 KMVYEDTLFLPFSGETVFMENPGLWILGCHNSDFRNGMTALLKVSSCDKNTGDYDE 739
Db 661 KMVYEDTLFLPFSGETVFMENPGLWILGCHNSDFRNGMTALLKVSSCDKNTGDYDE 720
QY 740 DSYEDISAYLLSKNNAIEPRSPQRSHRSTROKQFNATIPENDIEKTDPEWHAHTMP 799
Db 721 DSYEDISAYLLSKNNAIEPRSPQRSHRSTROKQFNATIPENDIEKTDPEWHAHTMP 780
QY 800 KIQNVSSDILLMLLRQSPTHGLSLDQAEKVFETSDPSGPAIDSNNSLSMTFRPQ 859
Db 781 KIQNVSSDILLMLLRQSPTHGLSLDQAEKVFETSDPSGPAIDSNNSLSMTFRPQ 840
QY 860 LHHSGMVFPPESGLQRLNEKLGTTAATLKKLDFKVSSTSNLJSTIPSDNLAAGTDN 919
Db 841 LHHSGMVFPPESGLQRLNEKLGTTAATLKKLDFKVSSTSNLJSTIPSDNLAAGTDN 900
QY 920 TSSIGPSPMPVHYDSQDITLFGKKSPLTESGGLSLSEENNDKLLBSGLNNSQESSW 979
Db 901 TSSIGPSPMPVHYDSQDITLFGKKSPLTESGGLSLSEENNDKLLBSGLNNSQESSW 960
QY 980 GKNVSTESGRLFKGRAGPALLTKDNALFKVSIISLLKTKNTSNKSNATNRKTHIDGPSL 1039
Db 961 GKNVSTESGRLFKGRAGPALLTKDNALFKVSIISLLKTKNTSNKSNATNRKTHIDGPSL 1020

QY 1040 LIENSPVQWQILSDTEFKKVTPLIHDRMLMDXNATRLNHNMSNKTSTSKNEMVQOK 1039
DB 1021 LIENSPVQWQILSDTEFKKVTPLIHDRMLMDXNATRLNHNMSNKTSTSKNEMVQOK 1030
QY 1100 KEGIPDDAQNPMDSFFKMLFLPSARWIQTHGKNSLNSCGGSPKOLVSLGPKSVG 1159
DB 1081 KEGIPDDAQNPMDSFFKMLFLPSARWIQTHGKNSLNSCGGSPKOLVSLGPKSVG 1140
QY 1160 QNFUSEKXNVVVGGEFTKDVGLKEMVPPSRNPLTLNLDLHNNTHNQSKKIQBELEK 1219
DB 1141 QNFUSEKXNVVVGGEFTKDVGLKEMVPPSRNPLTLNLDLHNNTHNQSKKIQBELEK 1200
QY 1220 KETLQENNVLPQIRTVGTGTNFMKNLFLSTRQNVESYGAYAPVQLQDPRSLNDSNR 1279
DB 1201 KETLQENNVLPQIRTVGTGTNFMKNLFLSTRQNVESYGAYAPVQLQDPRSLNDSNR 1260
QY 1280 TKKTAHESKGBEBNLEGLNQKQIYEKACTTRISPNTSQNFVQORSKRALQPRL 1339
DB 1261 TKKTAHESKGBEBNLEGLNQKQIYEKACTTRISPNTSQNFVQORSKRALQPRL 1320
QY 1340 PLETELEKRIIVDDTSTQWKNMKHLTPSLTQIDYNEKEKGATQSPLSDCLTRSHSI 1399
DB 1321 PLETELEKRIIVDDTSTQWKNMKHLTPSLTQIDYNEKEKGATQSPLSDCLTRSHSI 1380
QY 1400 POANRSPPIAKVSSFPPIRPIYLRVLFDQNSHLPAAASYRKDGSGVOESSHFJQGA 1459
DB 1381 POANRSPPIAKVSSFPPIRPIYLRVLFDQNSHLPAAASYRKDGSGVOESSHFJQGA 1440
QY 1460 NNLSLAULTLEMTGQREVSLGTSATNSVYKKVENTVLKPDLPKTSKGVKELLPKVHI 1519
DB 1441 NNLSLAULTLEMTGQREVSLGTSATNSVYKKVENTVLKPDLPKTSKGVKELLPKVHI 1500
QY 1520 YQKOLFPETNSNGSCHLDVEGSLQCTEGAIKWNEANRPCKVFLRVATESSAKTESK 1579
DB 1501 YQKOLFPETNSNGSCHLDVEGSLQCTEGAIKWNEANRPCKVFLRVATESSAKTESK 1560
QY 1580 LLDPLANDNHYGTQIPKEBWSQESPEKTAFAKKDITLSLNACSNHAIAINEGONKP 1639
DB 1561 LLDPLANDNHYGTQIPKEBWSQESPEKTAFAKKDITLSLNACSNHAIAINEGONKP 1620
QY 1640 ELEVTAQGRTERLCSONPPVLRKHOREIRFTTLOSQOEEIDYDDTISVEMKKECFDIY 1699
DB 1621 ELEVTAQGRTERLCSONPPVLRKHOREIRFTTLOSQOEEIDYDDTISVEMKKECFDIY 1680
QY 1700 DEDENQSPRSFQKTRHYFIAAVERLDWYGMSSPHVLNRAQSGSVQPKKVVQFEFTD 1759
DB 1681 DEDENQSPRSFQKTRHYFIAAVERLDWYGMSSPHVLNRAQSGSVQPKKVVQFEFTD 1740
QY 1760 GSFTQPLVYRGELNEHLGLGPIYRAEVEDNIMVTRNQASRPYSYSLISYEEQROGA 1819
DB 1741 GSFTQPLVYRGELNEHLGLGPIYRAEVEDNIMVTRNQASRPYSYSLISYEEQROGA 1800
QY 1820 BPRKNFVKNETKTYFWKVOHMAPTKDFDCKAWYPSDVLKDVHSLGLIGPLLVCHT 1879
DB 1801 BPRKNFVKNETKTYFWKVOHMAPTKDFDCKAWYPSDVLKDVHSLGLIGPLLVCHT 1860
QY 1880 NTLPNARGQVTVQBFALFTTIFDETKSWYFTENMERNACPCNIQMEDPTPKENYRPHA 1939
DB 1861 NTLPNARGQVTVQBFALFTTIFDETKSWYFTENMERNACPCNIQMEDPTPKENYRPHA 1920
QY 1940 INGYIMDTPLGLWMAQDQIRWYLLSMGNSNENIHSIFSGHYFTVRKKEEYQALYNLYP 1999
DB 1921 INGYIMDTPLGLWMAQDQIRWYLLSMGNSNENIHSIFSGHYFTVRKKEEYQALYNLYP 1980
QY 2000 GYFETVEMLPKAGIWRVCLGEHLHAGMSTLFLVYSNKCOTPLQMASGHIRDFOITAS 2059
DB 1981 GYFETVEMLPKAGIWRVCLGEHLHAGMSTLFLVYSNKCOTPLQMASGHIRDFOITAS 2040
QY 2060 GYGQWAPKLARIHYGSSINAWSTKEPPFSWIKVYDLIAPMIIGHIKTQGARQKPFSSLIYISQ 2119
DB 2041 GYGQWAPKLARIHYGSSINAWSTKEPPFSWIKVYDLIAPMIIGHIKTQGARQKPFSSLIYISQ 2100
QY 2120 FTIYSLDGKKWQTYRGNSTGTLMVFGNVDSSGIKHNIFNPPIARIYRLHPHYSIRS 2179

DB 2101 FTIYSLDGKKWQTYRGNSTGTLMVFGNVDSSGIKHNIFNPPIARIYRLHPHYSIRS 2160
QY 2180 TLRMELMGCDLNSCMPLGWSKAISSAQITASSYFTNMFATWSPSKARLHLQGRSNAWR 2239
DB 2161 TLRMELMGCDLNSCMPLGWSKAISSAQITASSYFTNMFATWSPSKARLHLQGRSNAWR 2220
QY 2240 PQVNNPKREWQVDPQKTMKVGTGTQGVKSLTSMYVKEFLISSQDGHQWTLFFQNGKV 2299
DB 2221 PQVNNPKREWQVDPQKTMKVGTGTQGVKSLTSMYVKEFLISSQDGHQWTLFFQNGKV 2280
QY 2300 KVFQGNQDSPTPVVNSLDPPLRLRYLRIHQSWVHQIALRMEVLGCEAODLY 2351
DB 2281 KVFQGNQDSPTPVVNSLDPPLRLRYLRIHQSWVHQIALRMEVLGCEAODLY 2332
RESULT 9
US-08-474-503-2
; Sequence 2, Application US/08474503
; Patent No. 574446
; GENERAL INFORMATION:
; APPLICANT: Emory University
; TITLE OF INVENTION: Hybrid Human/Animal Factor VIII
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Kilpatrick & Cody
; STREET: 1100 Peachtree Street, Suite 2800
; CITY: Atlanta
; STATE: Georgia
; COUNTRY: US
; ZIP: 30309
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/474,503
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Pratt, John S.
; REGISTRATION NUMBER: 29,476
; REFERENCE/DOCKET NUMBER: EMDU06CIP(3)
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 404-815-6500
; TELEFAX: 404-815-6555
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2332 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHEICAL: YES
; ANTI-SENSE: NO
; FRAGMENT TYPE: N-terminal
; ORIGINAL SOURCE:
; ORGANISM: Homo sapien
; TISSUE TYPE: Liver cDNA sequence
US-08-474-503-2
Query Match 96.9%; Score 12282; DB 1; Length 2332;
Best Local Similarity 95.8%; Pred. No. 0;
Matches 2327; Conservative 2; Mismatches 3; Indels 0; Gaps 0;
QY 20 ATRRYILGAVELSDYMQSDIGELPVDARFPFRPKSPPTNTSVYKTLFVEFDHLFN 79
DB 1 ATRRYILGAVELSDYMQSDIGELPVDARFPFRPKSPPTNTSVYKTLFVEFDHLFN 60
QY 80 IAKPRPPWMLGLGTQAEVYDVTITLKNASHPVSLHVGVSYWKASGAEYDDQTSQ 139
DB 61 IAKPRPPWMLGLGTQAEVYDVTITLKNASHPVSLHVGVSYWKASGAEYDDQTSQ 120

140 REKEDDXVFGGSHYYWVYKENGPMASDPLCLTYSYLSHVLDVKDLSGLGALLVCR 139
Db 121 REKEDDXVFGGSHYYWVYKENGPMASDPLCLTYSYLSHVLDVKDLSGLGALLVCR 180
Qy 200 EGSIAKEKTOTLHKFILLFAVDEGKSHWSETKNSLMQDRDAASARAWPKVHTVNGYVNR 259
Db 181 EGSIAKEKTOTLHKFILLFAVDEGKSHWSETKNSLMQDRDAASARAWPKVHTVNGYVNR 240
Qy 260 SLPLGLICHRKSVVWVIGVGTTPPEVHSIFLEGHTFLVRNHRQASLEISPIITLTAQTL 319
Db 241 SLPLGLICHRKSVVWVIGVGTTPPEVHSIFLEGHTFLVRNHRQASLEISPIITLTAQTL 300
Qy 320 MDLGOFLLSCHISHOHDGMEAYVYKVDSCPEEPOLINKNNEEAEDYDDDLTDSMDVVR 379
Db 301 MDLGOFLLSCHISHOHDGMEAYVYKVDSCPEEPOLINKNNEEAEDYDDDLTDSMDVVR 360
Qy 380 DDNSPSFQIRSVAKKHPXTVHYIAAEEEDMDYAPLVAPDDRYSKYSLNNGPQIRG 439
Db 361 DDNSPSFQIRSVAKKHPXTVHYIAAEEEDMDYAPLVAPDDRYSKYSLNNGPQIRG 420
Qy 440 RKYKVRPMAYTDTFTXTRAIOHESGILGPELLYCEVGDILLIIIFKQOASRPYNIYPHGI 499
Db 421 RKYKVRPMAYTDTFTXTRAIOHESGILGPELLYCEVGDILLIIIFKQOASRPYNIYPHGI 480
Qy 500 TDVREPLYSRRLPKGVKHLKOPPLPGBEFKYKWTVTVEDGPTKSDPRCLTRYSSFVNME 559
Db 481 TDVREPLYSRRLPKGVKHLKOPPLPGBEFKYKWTVTVEDGPTKSDPRCLTRYSSFVNME 540
Qy 560 RDLASGLIGPLLI CYKBSVDQKGNQIMSDKXENVILFSVDFENRSMYLTENIQRELNPAG 619
Db 541 RDLASGLIGPLLI CYKBSVDQKGNQIMSDKXENVILFSVDFENRSMYLTENIQRELNPAG 600
Qy 620 VQLEDPEFQASNIHMSINGYVFDLSQLSVCLHEVAYWYLLSIGAQOTPELSVFFSGYTFKH 679
Db 601 VQLEDPEFQASNIHMSINGYVFDLSQLSVCLHEVAYWYLLSIGAQOTPELSVFFSGYTFKH 660
Qy 680 KMVYEDTLTLPFSGEIVFNSMENPGWILGCHNSDFNRNMTALLKVSSCDKXNTGDYVE 739
Db 661 KMVYEDTLTLPFSGEIVFNSMENPGWILGCHNSDFNRNMTALLKVSSCDKXNTGDYVE 720
Qy 740 DSYEDISAYLLSKNAIIEPRFSQNSRHPSTROQFNATTIPENDIEKTDWFAHRTMP 799
Db 721 DSYEDISAYLLSKNAIIEPRFSQNSRHPSTROQFNATTIPENDIEKTDWFAHRTMP 780
Qy 800 KIQVSSSDLLMLRQSPTEPCLSLSDLOBAKYTFSDDPSPGALDSNNSLSETHFRPQ 859
Db 781 KIQVSSSDLLMLRQSPTEPCLSLSDLOBAKYTFSDDPSPGALDSNNSLSETHFRPQ 840
Qy 860 LHHSGDMVFTPESGLOLRNKLGTAAATELKLKLDKFVSSSTNNLI STIPSDNLAAGTDN 919
Db 841 LHHSGDMVFTPESGLOLRNKLGTAAATELKLKLDKFVSSSTNNLI STIPSDNLAAGTDN 900
Qy 920 TSSILGPPSMPVHYDQDITLLFGKKSPPTESGGPLSLSEENNDKLLLESGLMNSQBSW 979
Db 901 TSSILGPPSMPVHYDQDITLLFGKKSPPTESGGPLSLSEENNDKLLLESGLMNSQBSW 960
Qy 980 GKNVSSSTESGRLFKGKAHGPAJLTKONALFKVSI SLLKTKNTSNNASATNFKTHIDGPSL 1039
Db 961 GKNVSSSTESGRLFKGKAHGPAJLTKONALFKVSI SLLKTKNTSNNASATNFKTHIDGPSL 1020
Qy 1040 LIENSPPSVQWQILSDTEFFKVTPLIHDRMLMDKNATALRNHNMSKNTTSSKNMEMVQOK 1099
Db 1021 LIENSPPSVQWQILSDTEFFKVTPLIHDRMLMDKNATALRNHNMSKNTTSSKNMEMVQOK 1080
Qy 1100 KEGTIPDPAQNPDMSFFKMLFLPESARWQRTGHKNSLNSGQSPKOLVSLGPEKSVEG 1159
Db 1081 KEGTIPDPAQNPDMSFFKMLFLPESARWQRTGHKNSLNSGQSPKOLVSLGPEKSVEG 1140
Qy 1160 QNFILSEKNKVVVGKEFTKQVGLKEMVTPSSENLFLTNDLNLHNNTNQBKKIQEELK 1219
Db 1141 QNFILSEKNKVVVGKEFTKQVGLKEMVTPSSENLFLTNDLNLHNNTNQBKKIQEELK 1200

1220 KETLIQENVVLPQIHVTGTGKPMKNLFLLSRQNVESGYDGAYAPVLODFRSLNDSTNR 1279
Db 1201 KETLIQENVVLPQIHVTGTGKPMKNLFLLSRQNVESGYDGAYAPVLODFRSLNDSTNR 1260
Qy 1280 TKKHTAHFSPKGBEENLEGNGTKOIVERYACTTRISNTSQONFVTOQSKALQKFR 1339
Db 1261 TKKHTAHFSPKGBEENLEGNGTKOIVERYACTTRISNTSQONFVTOQSKALQKFR 1320
Qy 1340 PLEBTELEKXII VDDTSTQWSKNMKHLTSTLTQIDYNEKEKGAITQSPLSDCLTSHSI 1399
Db 1321 PLEBTELEKXII VDDTSTQWSKNMKHLTSTLTQIDYNEKEKGAITQSPLSDCLTSHSI 1380
Qy 1400 PQANRSPPLPIAKVSSPSPRIPIYLTIVLFCODNSSHLPAAASYRKDKSGVQBSSEHFLQAKK 1459
Db 1381 PQANRSPPLPIAKVSSPSPRIPIYLTIVLFCODNSSHLPAAASYRKDKSGVQBSSEHFLQAKK 1440
Qy 1460 NNLSLAILTLEMTGDQREVSGLSGT SATNSVTYKKVENTVL PKDPLPKTSKVKELLPKVHI 1519
Db 1441 NNLSLAILTLEMTGDQREVSGLSGT SATNSVTYKKVENTVL PKDPLPKTSKVKELLPKVHI 1500
Qy 1520 YOXDLPTETSNGPSGLHDVVEGSLQGTREGAIKWNENRPGKVPFLRVATESSAKTPSK 1579
Db 1501 YOXDLPTETSNGPSGLHDVVEGSLQGTREGAIKWNENRPGKVPFLRVATESSAKTPSK 1560
Qy 1580 LLDPLAWDNHYGTQIPKEBWKSOEKSPEKTAFAKKDITLSLNACESHAAIAAINEGONKP 1639
Db 1561 LLDPLAWDNHYGTQIPKEBWKSOEKSPEKTAFAKKDITLSLNACESHAAIAAINEGONKP 1620
Qy 1640 EIEVTWAKQRTERLCSQNPVVKRHOEITRTTLOSDQBEIDYDDTISVEMKEDDIY 1699
Db 1621 EIEVTWAKQRTERLCSQNPVVKRHOEITRTTLOSDQBEIDYDDTISVEMKEDDIY 1680
Qy 1700 DEBENQSPRSFOKTRHYFAAVERLWDYGMSSSPHVLNRASGSGVPEQKXVVOFTD 1759
Db 1681 DEBENQSPRSFOKTRHYFAAVERLWDYGMSSSPHVLNRASGSGVPEQKXVVOFTD 1740
Qy 1760 GSFTQPLYRGELNEHLGGLGPYIRAEVEDNIMVTFRNQASRPYSFYSSLI SYEBDQOQA 1819
Db 1741 GSFTQPLYRGELNEHLGGLGPYIRAEVEDNIMVTFRNQASRPYSFYSSLI SYEBDQOQA 1800
Qy 1820 EPRKNFKVKNETKTYFWKVOHHMAPTKDEDFCKAWAYFSDVDLEKDVHSLIGPLLCHT 1879
Db 1801 EPRKNFKVKNETKTYFWKVOHHMAPTKDEDFCKAWAYFSDVDLEKDVHSLIGPLLCHT 1860
Qy 1880 NTLNPAHGRQVTVQEPALFTTIFDETKSWYFTENMBNCRAPCNIO MEDTFFKENYRPHA 1939
Db 1861 NTLNPAHGRQVTVQEPALFTTIFDETKSWYFTENMBNCRAPCNIO MEDTFFKENYRPHA 1920
Qy 1940 INGYIMDTLEGLYMAQDQRI RWYLLSMGSNENIHSIFSGHVFTVRKKBYKMAVNLYP 1999
Db 1921 INGYIMDTLEGLYMAQDQRI RWYLLSMGSNENIHSIFSGHVFTVRKKBYKMAVNLYP 1980
Qy 2000 GVPETVEMLEPSKAGIWRVECLIGEHLAGHAGSTLFLVYSNKCOOTPLGNWASGHIRDFOITAS 2059
Db 1981 GVPETVEMLEPSKAGIWRVECLIGEHLAGHAGSTLFLVYSNKCOOTPLGNWASGHIRDFOITAS 2040
Qy 2060 GOYQOWAPKLARLHYSGSINAWSTKEPFSWKVDLLAPMI IHGIKTOGARQKFSLSYISQ 2119
Db 2041 GOYQOWAPKLARLHYSGSINAWSTKEPFSWKVDLLAPMI IHGIKTOGARQKFSLSYISQ 2100
Qy 2120 FIIMYSLDGKKWQTYRGNSTGTLMVFFGNVDSSGKINIPNPPIIARYIRLHPHYAIRS 2179
Db 2101 FIIMYSLDGKKWQTYRGNSTGTLMVFFGNVDSSGKINIPNPPIIARYIRLHPHYAIRS 2160
Qy 2180 TLRNMLGCDLNSCSMPLGMSKAI SDAQITASSYFTNVEATWSPSKARHLQGRSNAR 2239
Db 2161 TLRNMLGCDLNSCSMPLGMSKAI SDAQITASSYFTNVEATWSPSKARHLQGRSNAR 2220
Qy 2240 POWNNPKWLVQDPQKMKVGTGVTQGVKSLLSMYVKEFLISSQDGHQWTLFFQNGKV 2299
Db 2221 POWNNPKWLVQDPQKMKVGTGVTQGVKSLLSMYVKEFLISSQDGHQWTLFFQNGKV 2280
Qy 2300 KVFQGNQDSFTFVVNSLOPPLLYRLRIHFQSWVHQIALRXEVLGCEAQDLY 2351

2281 KVFQGNQDSFTEPVNSLDPLLTLYLRHPQSVWHQIALRMEVLGCEAQLDY 2332

```
Db 121 REKEDDKVFFGGSHYVWQVLKENGPMASDPLCLITYSLSHVDLVKDLNGLGALLVCR 180
QY 200 EGSLAKKKTOTLKFILLFAVDFEGKSWHSETKNSLMQDRDAASARAWPKWHTVNGYVNR 259
Db 181 EGSLAKKKTOTLKFILLFAVDFEGKSWHSETKNSLMQDRDAASARAWPKWHTVNGYVNR 240
QY 260 SLPLGLICHRKSVVHVIGMTTPEVHSIFLEGHTFLVRNHRQASLSISPIITLTAQTL 319
Db 241 SLPLGLICHRKSVVHVIGMTTPEVHSIFLEGHTFLVRNHRQASLSISPIITLTAQTL 300
QY 320 MDLCQFLLSCHISSHQHDCMEAYKVDSCPEEPOLIMKNNEAEADYDDDLTDSMDVVRP 379
Db 301 MDLCQFLLFCHISSHQHDCMEAYKVDSCPEEPOLRMKNNEAEADYDDDLTDSMDVVRP 360
QY 380 DDNSPSPIQIRSAVKKPKTWVHYIAAEEDWDYAPLVLAPDDRSYKSOYLNNGPQIRG 439
Db 361 DDNSPSPIQIRSAVKKPKTWVHYIAAEEDWDYAPLVLAPDDRSYKSOYLNNGPQIRG 420
QY 440 RYKVKVRPMAYTDETFKTREAIQHESGILGPLLYGEVGDITLLIIFKQASSPYNIYPHGI 499
Db 421 RYKVKVRPMAYTDETFKTREAIQHESGILGPLLYGEVGDITLLIIFKQASSPYNIYPHGI 480
QY 500 TDVRPLYSRRLPKGVKHLKOPPLPGEIFKYKWTVTVEDGPTKSDPRCLTRYYSFVWME 559
Db 481 TDVRPLYSRRLPKGVKHLKOPPLPGEIFKYKWTVTVEDGPTKSDPRCLTRYYSFVWME 540
QY 560 RDLASGLTGILLICYKESVDQKQIMSDKENVILFSVFDENRSMWYLPENIQRELPNPAG 619
Db 541 RDLASGLTGILLICYKESVDQKQIMSDKENVILFSVFDENRSMWYLPENIQRELPNPAG 600
QY 620 VQLEDPEFQASNIHNSINGYVFDLSQLSVCLHEVAYWYILSIGAQTDFLSVFGSYTFKH 679
Db 601 VQLEDPEFQASNIHNSINGYVFDLSQLSVCLHEVAYWYILSIGAQTDFLSVFGSYTFKH 660
QY 680 KMVYEDTLTLFPFGSETVFMSENPLGLWILGCHNSDPNRGWTALLKVSXCDKXGTGYE 739
Db 661 KMVYEDTLTLFPFGSETVFMSENPLGLWILGCHNSDPNRGWTALLKVSXCDKXGTGYE 720
QY 740 DSYEDISAYLLSKXNAIEPRSFQNSRHPSTROKQFNATTIPENDIEKTDWFAHRTMP 799
Db 721 DSYEDISAYLLSKXNAIEPRSFQNSRHPSTROKQFNATTIPENDIEKTDWFAHRTMP 780
QY 800 KIQVSSSDLLMLAROSPTPHGLSLSDLOEAKYETFDSPGALDSNNSSEWTHPRPQ 859
Db 781 KIQVSSSDLLMLAROSPTPHGLSLSDLOEAKYETFDSPGALDSNNSSEWTHPRPQ 840
QY 860 LHSXGDMVFTPESGLOLRNEKLGTTAATELKKLDPKVSSTSNLIPIPSDNLAAGTDN 919
Db 841 LHSXGDMVFTPESGLOLRNEKLGTTAATELKKLDPKVSSTSNLIPIPSDNLAAGTDN 900
QY 920 TSSLGPPSNPVHYDQSLDTTLFGKXSSPLTSGGPLSLSENNDSKLLSGLMNSQESSW 979
Db 901 TSSLGPPSNPVHYDQSLDTTLFGKXSSPLTSGGPLSLSENNDSKLLSGLMNSQESSW 960
QY 990 GKXVSTESGRILFKGKRAHGPALLTKONALFKVSIISLTKNTKSNNSATNKTHIDGSL 1039
Db 961 GKXVSTESGRILFKGKRAHGPALLTKONALFKVSIISLTKNTKSNNSATNKTHIDGSL 1020
QY 1040 LIENSFSPVWQNILESDETFKVTPLIHDRMLMDKNATALRNLHNSNKTSSKNMEMVQOK 1099
Db 1021 LIENSFSPVWQNILESDETFKVTPLIHDRMLMDKNATALRNLHNSNKTSSKNMEMVQOK 1080
QY 1100 KEGPIPPDAQNPDMSPFFKMLFLPESARWIQTHGHKNSLNSCGGSPKQVLGPEKSVEG 1159
Db 1081 KEGPIPPDAQNPDMSPFFKMLFLPESARWIQTHGHKNSLNSCGGSPKQVLGPEKSVEG 1140
QY 1160 QNFLEKKNKVVKGKFTKQVCLKEMVFPSSNPLFTNLMDLHNHNTNOKKIQEETEK 1219
Db 1141 QNFLEKKNKVVKGKFTKQVCLKEMVFPSSNPLFTNLMDLHNHNTNOKKIQEETEK 1200
QY 1220 KETLLQENNVLPQIHVTGTXNFMKNLFLSTPRQNVESYEGAYAPVLDQFRSLNDSNR 1279
Db 1201 KETLLQENNVLPQIHVTGTXNFMKNLFLSTPRQNVESYEGAYAPVLDQFRSLNDSNR 1260
```

RESULT 10

```
US-08-670-707A-2
; Sequence 2, Application US/08670707A
; Patent No. 5859204
; GENERAL INFORMATION:
; APPLICANT: Iollar, John S.
; TITLE OF INVENTION: Hybrid Human/Animal Factor VIII
; NUMBER OF SEQUENCES: 40
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Greenlee, Winner and Sullivan, P.C.
; STREET: 5370 Manhattan Circle Suite 201
; CITY: Boulder
; STATE: Colorado
; COUNTRY: USA
; ZIP: 80303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTX: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/670,707A
; FILING DATE: 26-JUN-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: WO PCT/US94/13200
; FILING DATE: 15-NOV-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/212,133
; FILING DATE: 11-MAR-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/864,004
; FILING DATE: 07-APR-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Greenlee, Lorraine L.
; REGISTRATION NUMBER: 27,894
; REFERENCE/DOCKET NUMBER: 75-95F
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 303/499-8080
; TELEFAX: 303/499-8089
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2332 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: not relevant
; MOLECULE TYPE: protein
; HYPOTHETICAL: YES
; ANTI-SENSE: NO
; FRAGMENT TYPE: N-terminal
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; TISSUE TYPE: Liver
US-08-670-707A-2

Query Match 98.9%; Score 12282; DB 2; Length 2332;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 2327; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 20 ATRRYYLGAVELSWDMQSDLGELPVDARPPRPVPSFPFN-SVYKTKLFEVFTDHLFN 7s
Db 1 ATRRYYLGAVELSWDMQSDLGELPVDARPPRPVPSFPFN-SVYKTKLFEVFTDHLFN 60
QY 80 IAKPRPPWMLGPTTQAEVYDVTWITLKNMASHPVSLHAGVSVYKASGEGAYDDQTSQ 139
Db 61 IAKPRPPWMLGPTTQAEVYDVTWITLKNMASHPVSLHAGVSVYKASGEGAYDDQTSQ 120
QY 140 REKEDDKVFFGGSHYVWQVLKENGPMASDPLCLITYSLSHVDLVKDLNGLGALLVCR 199
```

QY 1280 TKGHAFHFKGBBENLEGLNQTKQIIVEXYACTTRISPNSTQONFVQSKALKQFRL 1339
DB 1261 TKGHAFHFKGBBENLEGLNQTKQIIVEXYACTTRISPNSTQONFVQSKALKQFRL 1320
QY 1340 PLEETELEKRIIVDDTSTQSKWKHUTPSTLTQIDYNEKEKAITQSPSLSDCLTRHESI 1399
DB 1321 PLEETELEKRIIVDDTSTQSKWKHUTPSTLTQIDYNEKEKAITQSPSLSDCLTRHESI 1380
QY 1400 PQANRSLPIAKVSSFFSIRPIYILTRVLFDQNSHLPAAASYRKDQSGVQSSHFLOQAKK 1459
DB 1381 PQANRSLPIAKVSSFFSIRPIYILTRVLFDQNSHLPAAASYRKDQSGVQSSHFLOQAKK 1440
QY 1460 NNLSLAILTLEMTQDQREVQSLGTSATNSVTYKXVENTVLPKPDLPKTSKXVLELLPKVHI 1519
DB 1441 NNLSLAILTLEMTQDQREVQSLGTSATNSVTYKXVENTVLPKPDLPKTSKXVLELLPKVHI 1500
QY 1520 YQKDLFPETNSNGSPGHLDLVEGSLLOQTEGAIKWNENANPKVPFLRVATESSAKTFSK 1579
DB 1501 YQKDLFPETNSNGSPGHLDLVEGSLLOQTEGAIKWNENANPKVPFLRVATESSAKTFSK 1560
QY 1580 LLDPLAWNHYGTQIPKEWKSQBSPEKTAFAKKDTILSLNACSNHAIAAINEGQNK 1639
DB 1561 LLDPLAWNHYGTQIPKEWKSQBSPEKTAFAKKDTILSLNACSNHAIAAINEGQNK 1620
QY 1640 EIEVTWAKOQSTBELCSQNPVLKXHOBEITETTLQSDOBEIDYDDTISVEMKXEDFDIY 1699
DB 1621 EIEVTWAKOQSTBELCSQNPVLKXHOBEITETTLQSDOBEIDYDDTISVEMKXEDFDIY 1680
QY 1700 DEDNQSPRSFQKTRHYFIAAVERLDWYGMSSSPHVLNRQAQSGSVQPFKKVVFQFTD 1759
DB 1681 DEDNQSPRSFQKTRHYFIAAVERLDWYGMSSSPHVLNRQAQSGSVQPFKKVVFQFTD 1740
QY 1760 GSFTQPIYRGELNHELGLLGGYIRAEVEDNTMVTFRNQAQSPYFYSLSLISYBEDQOGA 1819
DB 1741 GSFTQPIYRGELNHELGLLGGYIRAEVEDNTMVTFRNQAQSPYFYSLSLISYBEDQOGA 1800
QY 1820 EPRCNFKZNETKYFKVVOHMAPTKDEPCKAWAYFSDVLEKDVHSGHIGLGLLVCHT 1879
DB 1801 EPRCNFKZNETKYFKVVOHMAPTKDEPCKAWAYFSDVLEKDVHSGHIGLGLLVCHT 1860
QY 1880 NTLNPAHGRQVTVQEFALFFIFDETBSWYFTENMERNCRAPCNQMEDPTFKENYRPHA 1939
DB 1861 NTLNPAHGRQVTVQEFALFFIFDETBSWYFTENMERNCRAPCNQMEDPTFKENYRPHA 1920
QY 1940 INGVIMDTLPGLVMAQDQRIWYLLSMGNSNIHSHFSGHVFTVRKKEEYKMALYNLYP 1999
DB 1921 INGVIMDTLPGLVMAQDQRIWYLLSMGNSNIHSHFSGHVFTVRKKEEYKMALYNLYP 1980
QY 2000 GVFEIVEMLPKAGIWRVECLIGELHAGMSTFLVYSNKQTPGLGMASGHIRDFQITAS 2059
DB 1981 GVFEIVEMLPKAGIWRVECLIGELHAGMSTFLVYSNKQTPGLGMASGHIRDFQITAS 2040
QY 2060 QYQGWAPKLARLHYSGSINAWSTKEPFSWIKVDLLAPMIIHGIKTOGARQKFSLSYISQ 2119
DB 2041 QYQGWAPKLARLHYSGSINAWSTKEPFSWIKVDLLAPMIIHGIKTOGARQKFSLSYISQ 2100
QY 2120 FIIMYSLDCKKWQTVRGNSTGTLVYFFGNVDSSGKKNIFPPPIIARVIRLHPHYSIRS 2179
DB 2101 FIIMYSLDCKKWQTVRGNSTGTLVYFFGNVDSSGKKNIFPPPIIARVIRLHPHYSIRS 2160
QY 2180 TLRMELMGCDLNSCSNPLQMSKALSDAQITASSYFTNNKATWSPSKARLHJQGSNAWR 2239
DB 2161 TLRMELMGCDLNSCSNPLQMSKALSDAQITASSYFTNNKATWSPSKARLHJQGSNAWR 2220
QY 2240 PQVNNPKWLQVDFQTKMTGVTITQYKSLTSMYVKEFLISSQDGHQWTLFPQNGKV 2299
DB 2221 PQVNNPKWLQVDFQTKMTGVTITQYKSLTSMYVKEFLISSQDGHQWTLFPQNGKV 2280
QY 2300 KYFQGNQDSFTPVVNSLDFPLLTRYLIRHPQSWHQAIALRMVLCGEAQDLY 2351
DB 2281 KYFQGNQDSFTPVVNSLDFPLLTRYLIRHPQSWHQAIALRMVLCGEAQDLY 2332

RESULT 11
US-09-037-601-2
; Sequence 2, Application US/09037601
; Patent No. 6180371
; GENERAL INFORMATION:
; APPLICANT: Lollar, John S.
; TITLE OF INVENTION: Hybrid Human/Animal Factor VIII
; NUMBER OF SEQUENCES: 40
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Greenlee, Winner and Sullivan, P.C.
; STREET: 5370 Manhattan Circle Suite 201
; CITY: Boulder
; STATE: Colorado
; COUNTRY: USA
; ZIP: 80303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/037,601
; FILING DATE: 26-JUN-1996
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: WO PCT/US94/13200
; FILING DATE: 15-NOV-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/212,133
; FILING DATE: 11-MAR-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/864,004
; FILING DATE: 07-APR-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Ferber, Donna M.
; REGISTRATION NUMBER: 33,878
; REFERENCE/DOCKET NUMBER: 75-95F
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 303/499-8080
; TELEFAX: 303/499-8089
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2332 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: not relevant
; MOLECULE TYPE: protein
; HYPOTHEetical: YES
; ANTI-SENSE: NO
; FRAGMENT TYPE: N-terminal
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; TISSUE TYPE: Liver
US-09-037-601-2

Query Match 98.9%; Score 12282; DB 3; Length 2332;
Best Local Similarity 99.8%; Pred. No 0;
Matches 2327; Conservative 2; Mismatches 3; Indels 0; Gaps 0;
QY 20 ATRRYILGAVLSWDYMQSDIGELPVDARPPRPVKSPFPNTSVVYKTLFVEFTDHLFN 79
DB 1 ATRRYILGAVLSWDYMQSDIGELPVDARPPRPVKSPFPNTSVVYKTLFVEFTVHLFN 60
QY 80 IAKPRPPMGLGPTIQAEVYDTVITLKNMASHPVSLHVGVSYWKASGAEYDDQTSQ 139
DB 61 IAKPRPPMGLGPTIQAEVYDTVITLKNMASHPVSLHVGVSYWKASGAEYDDQTSQ 120
QY 140 REKEDDKVFPQGSHTYVQVLKENGPMASDPLCLUTYSVLSHVDLVKDLNSGLIGALLVCR 199
DB 121 REKEDDKVFPQGSHTYVQVLKENGPMASDPLCLUTYSVLSHVDLVKDLNSGLIGALLVCR 180
QY 200 BGLSAKXKTQTLHKFILLFAVDFECKSWHSETKNSLQDRDAASARAWPKMHTVNGYNR 259

Db 181 EGSIAKXQTLHKFLLFAVDEGKSESETKNSLMOQDRDAASARAWPKXHTVNGYVR 240
Qy 260 SLPLGIGCHRSVYVHWITGMGTTPEVHISIFLGHTFLVNRHQASLEISPTIFLTAQTLL 319
Db 241 SLPLGIGCHRSVYVHWITGMGTTPEVHISIFLGHTFLVNRHQASLEISPTIFLTAQTLL 300
Qy 320 MDLGFLSLCHSSHCHDGMAYVAVDSCPEPOLIMKNNEAEEDYDDDLTDSMDVVR 379
Db 301 MDLGFLSLCHSSHCHDGMAYVAVDSCPEPOLIMKNNEAEEDYDDDLTDSMDVVR 360
Qy 380 DDNSPSEFIQSRVAKKHPTWVHYIAAEEDWDYAPLVLPDDPSYKSYQLNNGPQIG 439
Db 361 DDNSPSEFIQSRVAKKHPTWVHYIAAEEDWDYAPLVLPDDPSYKSYQLNNGPQIG 420
Qy 440 RYKXKVRMAYVDETFKTRAEATOHESGILGPLLYGEVGTLLIFPKQASRPYNIYPHGI 499
Db 421 RYKXKVRMAYVDETFKTRAEATOHESGILGPLLYGEVGTLLIFPKQASRPYNIYPHGI 480
Qy 500 TDVRPLYRRLPKGVKHLKDPPLPGEIPIKYKWTVTVEDGPTKSDPRCLTRYSSFVNM 559
Db 481 TDVRPLYRRLPKGVKHLKDPPLPGEIPIKYKWTVTVEDGPTKSDPRCLTRYSSFVNM 540
Qy 560 ZDLASGLIGPLLCYKESVDQGNQIMSDKRNVLIFSVPDENRWSYLNIQRFUNPAG 619
Db 541 ZDLASGLIGPLLCYKESVDQGNQIMSDKRNVLIFSVPDENRWSYLNIQRFUNPAG 600
Qy 620 VOLEDPEQASIMHSINGVYFDSIQLSVCLHEVAYWILSTCAOTDPLSVFFSGYTERH 679
Db 601 VOLEDPEQASIMHSINGVYFDSIQLSVCLHEVAYWILSTCAOTDPLSVFFSGYTERH 660
Qy 680 KMVYEDTTLTLPFSGETVMSMENPGLWILGCHNSDFNRGNTALLKYSSCDKNTGDYIE 739
Db 661 KMVYEDTTLTLPFSGETVMSMENPGLWILGCHNSDFNRGNTALLKYSSCDKNTGDYIE 720
Qy 740 DSYEDISAYLLSKNAIBPSFQNSRHPSSTKQKFNATTIPENDIEKTDPMFAHRTMP 799
Db 721 DSYEDISAYLLSKNAIBPSFQNSRHPSSTKQKFNATTIPENDIEKTDPMFAHRTMP 780
Qy 800 KIONVSSDLMMLRQSPHGLSLSDLOEAKYETPSDDPSGATDKNNSLSEMTHERPQ 859
Db 781 KIONVSSDLMMLRQSPHGLSLSDLOEAKYETPSDDPSGATDKNNSLSEMTHERPQ 840
Qy 860 LHSXGDMVFTPESGQLRINEKLGTTAATLKKLDFKYSSTSNLITIPSDNLAAGTDN 919
Db 841 LHSXGDMVFTPESGQLRINEKLGTTAATLKKLDFKYSSTSNLITIPSDNLAAGTDN 900
Qy 920 TSSLGPPSMVHYDQOLDTTLPGKXSSPLTESGGPLSISEENNDSKLIESGLMNSQESSW 979
Db 901 TSSLGPPSMVHYDQOLDTTLPGKXSSPLTESGGPLSISEENNDSKLIESGLMNSQESSW 960
Qy 980 GKNVSTSGRLFKGRAGPALLTKDNALFKVYSISLLTKNTKNSATNRKTHIDGPSL 1039
Db 961 GKNVSTSGRLFKGRAGPALLTKDNALFKVYSISLLTKNTKNSATNRKTHIDGPSL 1020
Qy 1040 LIENSFVWQNLIESTEPKVTPLIHDNMLDKNATLRLNHNKNTTSSKNMVMVQOK 1099
Db 1021 LIENSFVWQNLIESTEPKVTPLIHDNMLDKNATLRLNHNKNTTSSKNMVMVQOK 1080
Qy 1100 KEGPIPPDACNPMSEFFKMLFIPESARWIORTHGKNSLNSGOGSPKQVLSIGPEKSV 1159
Db 1081 KEGPIPPDACNPMSEFFKMLFIPESARWIORTHGKNSLNSGOGSPKQVLSIGPEKSV 1140
Qy 1160 QNFLSEKKNVYVKGKFTKDVGLKENVFPSSRNLFJTNLNHNHNTNHNQEKKIOEEIEK 1219
Db 1141 QNFLSEKKNVYVKGKFTKDVGLKENVFPSSRNLFJTNLNHNHNTNHNQEKKIOEEIEK 1200
Qy 1220 KETLIQENVVLPOIHTVTGKNFMKNLFIILSTRQNEVSGYDAGYAPVLOPFRSLNDSTNR 1279
Db 1201 KETLIQENVVLPOIHTVTGKNFMKNLFIILSTRQNEVSGYDAGYAPVLOPFRSLNDSTNR 1260
Qy 1280 TKKHTAFSKKGEENLEGLGNQTKQIWEKYACTTRISPNTSQONFVTQSRKALKOPRL 1339

RESULT 12

US-09-315-179-2

Db 1261 TKKHTAFSKKGEENLEGLGNQTKQIWEKYACTTRISPNTSQONFVTQSRKALKOPRL 1320
Qy 1340 PLEETELEKRIIVDDTSTQWSKNKHLTPSLTQIDYNEKEKGAITOSPISDCLTRSHSI 1399
Db 1321 PLEETELEKRIIVDDTSTQWSKNKHLTPSLTQIDYNEKEKGAITOSPISDCLTRSHSI 1380
Qy 1400 POANRSLPIAKVSSFPISIRPIYILTRVLFDONSSHLPAAASYRKXKDSGVQESSHFLQAKK 1459
Db 1381 POANRSLPIAKVSSFPISIRPIYILTRVLFDONSSHLPAAASYRKXKDSGVQESSHFLQAKK 1440
Qy 1460 NNLISALITLMTQDQREBVSGLGTSATNSVTYKVENIVLPKPDLPKTSKGVELLPKVI 1519
Db 1441 NNLISALITLMTQDQREBVSGLGTSATNSVTYKVENIVLPKPDLPKTSKGVELLPKVI 1500
Qy 1520 YOKLFPETISNGSPGHLDLVEGSLLOGTGCAIKWNEANRPKVPFFLRVATESAKTPSK 1579
Db 1501 YOKLFPETISNGSPGHLDLVEGSLLOGTGCAIKWNEANRPKVPFFLRVATESAKTPSK 1560
Qy 1580 LLDPLANDNHYGTQIPKEEWSQEKSPKTAFAKKKDTILSLNACESNHAIAAINEGQNK 1639
Db 1561 LLDPLANDNHYGTQIPKEEWSQEKSPKTAFAKKKDTILSLNACESNHAIAAINEGQNK 1620
Qy 1640 BIEVTWAKQGRTEBLCSQNPVILAKHOREITRITLQSDOEBIDYDDTISVEMKEDFDIY 1699
Db 1621 BIEVTWAKQGRTEBLCSQNPVILAKHOREITRITLQSDOEBIDYDDTISVEMKEDFDIY 1680
Qy 1700 DEDENQSPRSFQKTRHYFIAAVERLWDYGMSSPHVLNRNAQSGSVPOPKVVFQEPD 1759
Db 1681 DEDENQSPRSFQKTRHYFIAAVERLWDYGMSSPHVLNRNAQSGSVPOPKVVFQEPD 1740
Qy 1760 GSFTQPLVYRGLNEHLGLGPYIRAEVEDNIMVTFRNQASRPYFYSLSIYESDQROGA 1819
Db 1741 GSFTQPLVYRGLNEHLGLGPYIRAEVEDNIMVTFRNQASRPYFYSLSIYESDQROGA 1800
Qy 1820 EPRKNFKPNETKTYFWKVOHHMPTKDEPCKAWAYFSDVDLEKDVHSGILGPLVCHT 1879
Db 1801 EPRKNFKPNETKTYFWKVOHHMPTKDEPCKAWAYFSDVDLEKDVHSGILGPLVCHT 1860
Qy 1880 NTLNPAHQGVTVCEFPALFTTIFDETKSWYFTENWERNCRAPCNIQMEDPTFKENYRPHA 1939
Db 1861 NTLNPAHQGVTVCEFPALFTTIFDETKSWYFTENWERNCRAPCNIQMEDPTFKENYRPHA 1920
Qy 1940 INGYIMDTLPGLVAAQDQRIWYLLSMGSENIHSHPSGHVFTVRKKEBYKMALYNLYP 1999
Db 1921 INGYIMDTLPGLVAAQDQRIWYLLSMGSENIHSHPSGHVFTVRKKEBYKMALYNLYP 1980
Qy 2000 GVFEIVEMLPKAGIWRVECLIGELHAGMSTFLVYNSKQCTPLGMAAGHIRDFOITAS 2059
Db 1981 GVFEIVEMLPKAGIWRVECLIGELHAGMSTFLVYNSKQCTPLGMAAGHIRDFOITAS 2040
Qy 2060 GOYGOWAPKLARLEYSGSINAWSTKEPFSWIKVDLLAPMIHGIKTQARQKFSLSYISQ 2119
Db 2041 GOYGOWAPKLARLEYSGSINAWSTKEPFSWIKVDLLAPMIHGIKTQARQKFSLSYISQ 2100
Qy 2120 FIIMYSLDGKKWOYIRGNSGTGLMVFFGNVDSSGIKHNIHNPPIIARIYIRLHPHYSIES 2179
Db 2101 FIIMYSLDGKKWOYIRGNSGTGLMVFFGNVDSSGIKHNIHNPPIIARIYIRLHPHYSIES 2160
Qy 2180 TURMELMGCDLNSCNPGLMESKALISDAQITASSYFTNNFATWSPSKARLHLOGRSNWR 2239
Db 2161 TURMELMGCDLNSCNPGLMESKALISDAQITASSYFTNNFATWSPSKARLHLOGRSNWR 2220
Qy 2240 POWNPKEWLQVDFOKTMKVTVGTTTQGVKSLTSMYVKEPLISSQDQCHQWTLFPQNGKV 2299
Db 2221 POWNPKEWLQVDFOKTMKVTVGTTTQGVKSLTSMYVKEPLISSQDQCHQWTLFPQNGKV 2280
Qy 2300 KYFQGNQDSFTPVNSLDPPILTRIRTHPOSWHQIALRMEVLGCEAQDLY 2351
Db 2281 KYFQGNQDSFTPVNSLDPPILTRIRTHPOSWHQIALRMEVLGCEAQDLY 2332

Db 1141 QNFISEKKNVVGKGTVDGLKEMVFPSSRLFTJNLNLHNNTHNQEKIQEIEK 1200
QY 1220 KETLIQBNVVLPOHVTGTCIKNFKNLFLSTQNVEGSDGAYAPVLQDFRSLNDSTNR 1279
Db 1201 KETLIQBNVVLPOHVTGTCIKNFKNLFLSTQNVEGSDGAYAPVLQDFRSLNDSTNR 1260
QY 1280 TKKHTAHFSKKGHEENLEGGNOTKQIVKXACTTRISPNSTQONFVQTSKALKQFRL 1339
Db 1261 TKKHTAHFSKKGHEENLEGGNOTKQIVKXACTTRISPNSTQONFVQTSKALKQFRL 1320
QY 1340 PLEBETELEKRIIIVDDTSTQWKNMKHLTPSLTQIDYNEKEKGALITQSPISDCLTRSHSI 1399
Db 1321 PLEBETELEKRIIIVDDTSTQWKNMKHLTPSLTQIDYNEKEKGALITQSPISDCLTRSHSI 1380
QY 1400 POANRSLPFAKVSSPFSIRPIYILTRVLFDQNSHLPAAASYRKDSGVQESSHPLOAKK 1459
Db 1381 POANRSLPFAKVSSPFSIRPIYILTRVLFDQNSHLPAAASYRKDSGVQESSHPLOAKK 1440
QY 1460 NNLSLAILTLEMTGDQREVGLSTSATNSVTYKXVENTVLPKPDLPKTSQKVELLPKVHI 1519
Db 1441 NNLSLAILTLEMTGDQREVGLSTSATNSVTYKXVENTVLPKPDLPKTSQKVELLPKVHI 1500
QY 1520 YQKDLFPETSNQSGFHLDLVEGSLLOCTEGAIKXNEANRPGKVPFLRVATESSAKTPSK 1579
Db 1501 YQKDLFPETSNQSGFHLDLVEGSLLOCTEGAIKXNEANRPGKVPFLRVATESSAKTPSK 1560
QY 1580 LLDPLANDNHYGTQIPKBEWKSQSKSEKTAFAKKDTILSNACESHAAIAINEGONKP 1639
Db 1561 LLDPLANDNHYGTQIPKBEWKSQSKSEKTAFAKKDTILSNACESHAAIAINEGONKP 1620
QY 1640 EIEVTWAKQGTERTLCQNPPLVLRHOREIRTRTLQSDQEBIDYDDTISVEMKKEDFDIY 1699
Db 1621 EIEVTWAKQGTERTLCQNPPLVLRHOREIRTRTLQSDQEBIDYDDTISVEMKKEDFDIY 1680
QY 1700 DEDNQSPRSQKTRIFYTAAVRLMDYGMSSSPHVLNRNQAQSGVPQPKVYVQFPTD 1759
Db 1681 DEDNQSPRSQKTRIFYTAAVRLMDYGMSSSPHVLNRNQAQSGVPQPKVYVQFPTD 1740
QY 1760 GSFTQPLRYGELNEHLLGLGYIRAEVEDNIMVTRNQAQSGVPQPKVYVQFPTD 1819
Db 1741 GSFTQPLRYGELNEHLLGLGYIRAEVEDNIMVTRNQAQSGVPQPKVYVQFPTD 1800
QY 1820 EPRKNFVKPNETKTYFWKVVQHMAPTKDEPCKAWAYFSDVLEKDVHSGLGIGLLVCHT 1879
Db 1801 EPRKNFVKPNETKTYFWKVVQHMAPTKDEPCKAWAYFSDVLEKDVHSGLGIGLLVCHT 1860
QY 1880 NTLNPAHGRQVTVQEPALFFTIIDETKSWYFTENMERNCRAPCNIQMEDDPTFKENYRPHA 1939
Db 1861 NTLNPAHGRQVTVQEPALFFTIIDETKSWYFTENMERNCRAPCNIQMEDDPTFKENYRPHA 1920
QY 1940 INGVIMDTPLGLVMAQDQIRIRWYLLSMGNSNIHSHFSGHVTVVRKXEYKMALYNLYP 1999
Db 1921 INGVIMDTPLGLVMAQDQIRIRWYLLSMGNSNIHSHFSGHVTVVRKXEYKMALYNLYP 1980
QY 2000 GVFTVEMLPKAGIRWVECHLGHAGMTLFLVYGNKQOTPLGMASGHIRDFQITAS 2059
Db 1981 GVFTVEMLPKAGIRWVECHLGHAGMTLFLVYGNKQOTPLGMASGHIRDFQITAS 2040
QY 2060 GCYQWAPKLARLHYSGSINAWSKPEPSWIKVDLLAPMIHGIKTQAROKFSSLYISQ 2119
Db 2041 GCYQWAPKLARLHYSGSINAWSKPEPSWIKVDLLAPMIHGIKTQAROKFSSLYISQ 2100
QY 2120 FIIMYSLOGKWQTVRGNSTGFLMYFPGVDSGGIKHNIENPPIIARYIRLHPHYSIRS 2179
Db 2101 FIIMYSLOGKWQTVRGNSTGFLMYFPGVDSGGIKHNIENPPIIARYIRLHPHYSIRS 2160
QY 2180 TLRMLMGCDLNSCSMPLGMSKAIISDAQITASSYFTNNFATWSPSKARLHLQGSNAWR 2239
Db 2161 TLRMLMGCDLNSCSMPLGMSKAIISDAQITASSYFTNNFATWSPSKARLHLQGSNAWR 2220
QY 2240 PQVNNPKWLQVDFQKTMKVTVGTQGVKSLTSTMYVKEFLISSQDQHQWTLFFQNGKV 2299

Db 2221 PQVNNPKWLQVDFQKTMKVTVGTQGVKSLTSTMYVKEFLISSQDQHQWTLFFQNGKV 2280
QY 2300 KVFQGNQDSTPFWNSLDPPLLTRYLRIHQSWVHQIALAMEVLGCEAODLY 2351
Db 2281 KVFQGNQDSTPFWNSLDPPLLTRYLRIHQSWVHQIALAMEVLGCEAODLY 2332
RESULT 14
PCT-US93-03275-4
; Sequence 4, Application PC/TUS9303275
; GENERAL INFORMATION:
; APPLICANT: Lollar, John S.
; APPLICANT: Ruge, Marshall S.
; TITLE OF INVENTION: Hybrid Human/Porcine Factor VIII
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Kilpatrick & Cody
; STREET: 1100 Peachtree Street, Suite 2800
; CITY: Atlanta
; STATE: Georgia
; COUNTRY: US
; ZIP: 30309-4530
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent-In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US93/03275
; FILING DATE: 19930407
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/864004
; FILING DATE: 07-APR-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Pabst, Patrea L.
; REGISTRATION NUMBER: 31,284
; REFERENCE/DOCKET NUMBER: BMU 106PCT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 404-815-6508
; TELEFAX: 404-815-6555
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2332 amino acids
; TYPE: AMINO ACID
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHEICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: N-terminal
; ORIGINAL SOURCE:
; ORGANISM: Homo sapien
; TISSUE TYPE: Liver cdna sequence
PCT-US93-03275-4
Query Match 98.9%; Score 12282; DB 5; Length 2332;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 2327; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 20 ATRRYIYGAVELSDWYQSDIGELPVDARPPRPVKSPFNWTSVYKTLFVETDHLFN 79
Db 1 ATRRYIYGAVELSDWYQSDIGELPVDARPPRPVKSPFNWTSVYKTLFVETDHLFN 60
QY 80 IAKRPPPMWGLLGTIOAEVYDVTVTILKNASHPVSLHAGVSYWKASGAEYDDQTSQ 139
Db 61 IAKRPPPMWGLLGTIOAEVYDVTVTILKNASHPVSLHAGVSYWKASGAEYDDQTSQ 120
QY 140 REKEDDKVFGGSHYVYQVILKENGPMASPLCLITYSLSHVDLVKDLNSGLIGALLVCR 199
Db 121 REKEDDKVFGGSHYVYQVILKENGPMASPLCLITYSLSHVDLVKDLNSGLIGALLVCR 180
QY 200 EGSIAKERTQTLHKFILLFAVFDGKSHSETKNSLMQDRDAASARAWPKMHTVNGYVNR 259

181 EGS LAKE KNTQTLHKFILLFANVDFCKSWHSETKNSLXQORDAASARAPKWHYVNGYNR 240
260 SLPLGLIGHCRKSVYVHWITGMGTTTPPVHISIFLEGHTFLVNRHQASLEISPTTFITTAQTL 319
241 SLPLGLIGHCRKSVYVHWITGMGTTTPPVHISIFLEGHTFLVNRHQASLEISPTTFITTAQTL 300
320 MDLQGFLLSCHTSSHQHGMAYVVDSCPEPQLIMKNNEAEADYDDDLTDSMDVVR 379
301 MDLQGFLLFCHTSSHQHGMAYVVDSCPEPQLIMKNNEAEADYDDDLTDSMDVVR 360
380 DDNSPSPQIRSVAKKHPKTVWHYIAAEEDWDYAPLVLPADDDRSYKSOYLNNGPQRIG 439
361 DDNSPSPQIRSVAKKHPKTVWHYIAAEEDWDYAPLVLPADDDRSYKSOYLNNGPQRIG 420
440 RYKVKVRNAYTDETFKTRREALQHESGILGPLLYGEVGETLLIIFKNQASRPYNIYPHGI 499
421 RYKVKVRNAYTDETFKTRREALQHESGILGPLLYGEVGETLLIIFKNQASRPYNIYPHGI 480
500 TVRPLYRRPLPKGVKHLKDPPLPGETPFYKWTVTVEDGPTKSDPRCLTRYSSFVNM 559
481 TVRPLYRRPLPKGVKHLKDPPLPGETPFYKWTVTVEDGPTKSDPRCLTRYSSFVNM 540
560 RDLASGLIGPLLICYKESVDQGNQIMSDKRNVLFSVFCNRSWYLATENIQRELNPAG 619
541 RDLASGLIGPLLICYKESVDQGNQIMSDKRNVLFSVFCNRSWYLATENIQRELNPAG 600
620 VOLEDPRQASNMHSINGVYFDSIOLSVCLHEVAYWYILSIGAQDTFLSFYSSYTFKH 679
601 VOLEDPRQASNMHSINGVYFDSIOLSVCLHEVAYWYILSIGAQDTFLSFYSSYTFKH 660
680 KMVYEDTTLTLPFSGETVPMSENDEGLMILGCHNSDFNRGWTALLKYSSODKNTGDYYE 739
661 KMVYEDTTLTLPFSGETVPMSENDEGLMILGCHNSDFNRGWTALLKYSSODKNTGDYYE 720
740 DSYEDISAYLSKNNAISEPRFSQNSRHPSTRQKQFNATTIPENDIEKTDPFWAHRTMP 799
721 DSYEDISAYLSKNNAISEPRFSQNSRHPSTRQKQFNATTIPENDIEKTDPFWAHRTMP 780
800 KIONVSSDLMALRQSTPHGLSLSDIQEAKYETFSDDPGALDSNNISEMTHFRPQ 859
781 KIONVSSDLMALRQSTPHGLSLSDIQEAKYETFSDDPGALDSNNISEMTHFRPQ 840
860 LHSQDMVFTPESGIQLRINKLGTAAATELKKLDFKVSSTSNLI STIPSDNLAAGTDN 919
841 LHSQDMVFTPESGIQLRINKLGTAAATELKKLDFKVSSTSNLI STIPSDNLAAGTDN 900
920 TSSLGPPMPVHYDSQDITTLFGKKSPLTESGGPLSISEENNDKLLSGLMNSQESSW 979
901 TSSLGPPMPVHYDSQDITTLFGKKSPLTESGGPLSISEENNDKLLSGLMNSQESSW 960
980 GKNVSSITESGRLFCKGRAGPALLTKDNALFKVSI SLLKTKNTKSNNSATNRKTHIDGPSL 1039
961 GKNVSSITESGRLFCKGRAGPALLTKDNALFKVSI SLLKTKNTKSNNSATNRKTHIDGPSL 1020
1040 LIENSPPYQWNILEDTEBFKVTPLIHRMLMDKNATALRLNEMGNKTTSSKNMEMWQOK 1099
1021 LIENSPPYQWNILEDTEBFKVTPLIHRMLMDKNATALRLNEMGNKTTSSKNMEMWQOK 1080
1100 KEGPIPPDAQNDMPFFKMLFLPESARWTQTHGKNLSNGGQSPKQIVSLGPSPKSV 1159
1081 KEGPIPPDAQNDMPFFKMLFLPESARWTQTHGKNLSNGGQSPKQIVSLGPSPKSV 1140
1160 QNFLSEKXKVVVGKEFTKDVGLKEMVFPSSNLFILTNLDNLHENNTHNCKKIQEEIEK 1219
1141 QNFLSEKXKVVVGKEFTKDVGLKEMVFPSSNLFILTNLDNLHENNTHNCKKIQEEIEK 1200
1220 KETLIQENYVLPQIHTVTGKNFMKNLFLLSRQNVESYDGAAPVQDPRSLNDSNR 1279
1201 KETLIQENYVLPQIHTVTGKNFMKNLFLLSRQNVESYDGAAPVQDPRSLNDSNR 1260
1280 TKKHTAHSKKEEENLEGLNQTKQIVEKYACTTRISPNSTSQNQFVTRQSRKALKQFRL 1339

1261 TKKHTAHSKKEEENLEGLNQTKQIVEKYACTTRISPNSTSQNQFVTRQSRKALKQFRL 1320
1340 PLEETBELKRIIIVDDTSTOWSKNMKHLTPSTLTQIDYNEKEKGALTQSPISDCLTRHSHI 1399
1321 PLEETBELKRIIIVDDTSTOWSKNMKHLTPSTLTQIDYNEKEKGALTQSPISDCLTRHSHI 1380
1400 PQANRSPPLAKVSSFFSIRPIYILTRVLFDQNSHLPAAASYRKXKDSGVQZESSHFLQGAKK 1459
1381 PQANRSPPLAKVSSFFSIRPIYILTRVLFDQNSHLPAAASYRKXKDSGVQZESSHFLQGAKK 1440
1460 NNLSLAILTLEMTQDQREVSGLGTSATNSVTYKXVENTVLPKPLPKTSGKVELLPKVHI 1519
1441 NNLSLAILTLEMTQDQREVSGLGTSATNSVTYKXVENTVLPKPLPKTSGKVELLPKVHI 1500
1520 YQXDLFPTETSNQSPGHLDLVEGSLLOGTGCAIKWNEANRPGKVPFLRVATESSAKTPSK 1579
1501 YQXDLFPTETSNQSPGHLDLVEGSLLOGTGCAIKWNEANRPGKVPFLRVATESSAKTPSK 1560
1580 LLDPLANDNHVGTQIPKEEWKSOEKSPEKTAFAKKDITLSLNACESNHAIAAINEGQNK 1639
1561 LLDPLANDNHVGTQIPKEEWKSOEKSPEKTAFAKKDITLSLNACESNHAIAAINEGQNK 1620
1640 EIEVTWAKQGTRELCSQNPVLPKXHORBITRTTLOSQOBBIDYDDTISIVEMKKEDEIY 1699
1621 EIEVTWAKQGTRELCSQNPVLPKXHORBITRTTLOSQOBBIDYDDTISIVEMKKEDEIY 1680
1700 DEENQSPRSCFKKTRHYFTAAVERLWDYGNSSPHVLNKAQSGSVPOKXVYQBEFTD 1759
1681 DEENQSPRSCFKKTRHYFTAAVERLWDYGNSSPHVLNKAQSGSVPOKXVYQBEFTD 1740
1760 GSFTQPLRYRGLNBHGLGLLPYIRAEVEDNIMVTRFQAQRPYSFYSSLSYBBDQOQGA 1819
1741 GSFTQPLRYRGLNBHGLGLLPYIRAEVEDNIMVTRFQAQRPYSFYSSLSYBBDQOQGA 1800
1820 BPRXNFVKNETKTYFKVQVHHMAPTKDEDFCKAWAYFSDVLEKDVHSGILGELLVCHT 1879
1801 BPRXNFVKNETKTYFKVQVHHMAPTKDEDFCKAWAYFSDVLEKDVHSGILGELLVCHT 1860
1880 NTLAPAGROVTVQEFALFTTIDETKSWPTENNERNCRAPCNIOMEDPTFKENYRPHA 1939
1861 NTLAPAGROVTVQEFALFTTIDETKSWPTENNERNCRAPCNIOMEDPTFKENYRPHA 1920
1940 INGVIMDTPLGLVMAQDQIRWYLLSNGSNIHISIFSGHVFTVRKKEEYKMALYNLYP 1999
1921 INGVIMDTPLGLVMAQDQIRWYLLSNGSNIHISIFSGHVFTVRKKEEYKMALYNLYP 1980
2000 GVFETVEMLPKAGIWRVECLIGEHLFAGMSTLPLVYSNKCOTPLGKASGHIRDFQITAS 2059
1981 GVFETVEMLPKAGIWRVECLIGEHLFAGMSTLPLVYSNKCOTPLGKASGHIRDFQITAS 2040
2060 QOYGOWAPKLARLHYSGSINAWSTKEPFSWIKVDLLAPMIHGIKTQAROKPSSLVYSQ 2119
2041 QOYGOWAPKLARLHYSGSINAWSTKEPFSWIKVDLLAPMIHGIKTQAROKPSSLVYSQ 2100
2120 FIIMYSLDGKQWQYRGNSTGTLMVFFGNVDSSGKINFNPPIIARYIRLHPHTHSIRS 2179
2101 FIIMYSLDGKQWQYRGNSTGTLMVFFGNVDSSGKINFNPPIIARYIRLHPHTHSIRS 2160
2180 TLRVLMGCDLNSCSMPLGMSKASIDAOITASSYFTNMFATWSPSKARLHQESNAWR 2239
2161 TLRVLMGCDLNSCSMPLGMSKASIDAOITASSYFTNMFATWSPSKARLHQESNAWR 2220
2240 PQVNNPEWLOVDQKTMKVTVGTQGVKSLTSMYVKEFLISSSQDGHOWTLFPQNGKV 2299
2221 PQVNNPEWLOVDQKTMKVTVGTQGVKSLTSMYVKEFLISSSQDGHOWTLFPQNGKV 2280
2300 KVFQGNQDSFPPVNSLDPPLLTTRYLRIHPQSWVHQIALRMEVLGCEAQDLY 2351
2281 KVFQGNQDSFPPVNSLDPPLLTTRYLRIHPQSWVHQIALRMEVLGCEAQDLY 2332

```

; Sequence 2, Application PC/TUS9413200
; GENERAL INFORMATION:
; APPLICANT: Emory University
; TITLE OF INVENTION: Hybrid Human/Animal Factor VIII
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Kilpatrick & Cody
; STREET: 1100 Peachtree Street, Suite 2800
; CITY: Atlanta
; STATE: Georgia
; COUNTRY: US
; ZIP: 30309
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US94/13200
; FILING DATE: 15-NOV-1994
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Pabst, Patrea L.
; REGISTRATION NUMBER: 31,284
; REFERENCE/DOCKET NUMBER: EMU106CIP(2)
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 404-815-6508
; TELEFAX: 404-815-6555
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2332 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: YES
; ANTI-SENSE: NO
; FRAGMENT TYPE: N-terminal
; ORIGINAL SOURCE:
; ORGANISM: Homo sapien
; TISSUE TYPE: Liver cDNA sequence
; PCT-US94-13200-2

Query Match 98.9%; Score 12282; DB 5; Length 2332;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 2327; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Qy 20 ATRRYILGAVELSDYMQSDLGELPVDARPPRPVKGFPPFNTSVVYKTLFVEFTLHFN 79
Db 1 ATRRYILGAVELSDYMQSDLGELPVDARPPRPVKGFPPFNTSVVYKTLFVEFTLHFN 60

Qy 80 IAKPRPMGLGPTIOAEVYDVTVITLKNASHFVSLHVGVSVMKASGEGAYDDQTSQ 139
Db 61 IAKPRPMGLGPTIOAEVYDVTVITLKNASHFVSLHVGVSVMKASGEGAYDDQTSQ 120

Qy 140 REKDDKVFPGSGSYVYQVILKENGPMASDPLCLITYSLSHVDVLKQINSGLIGALLVCR 199
Db 121 REKDDKVFPGSGSYVYQVILKENGPMASDPLCLITYSLSHVDVLKQINSGLIGALLVCR 180

Qy 200 EGSIAKERTQTLHKPILLFAVFDGKSWHSTKNSLMQDRDAASARAWPKKHTYNGYNR 259
Db 181 EGSIAKERTQTLHKPILLFAVFDGKSWHSTKNSLMQDRDAASARAWPKKHTYNGYNR 240

Qy 260 SLPLGIGCHRSVYWHVIGMGTTEVEVHSIFLEGHTFLVRNHRQASLEISPTIFLTAQTL 319
Db 241 SLPLGIGCHRSVYWHVIGMGTTEVEVHSIFLEGHTFLVRNHRQASLEISPTIFLTAQTL 300

Qy 320 MDLQCFLLSHSHQHDGMAYVKVDCSCEPQOLIMQNEEASDYDDLDSMDVVRP 379
Db 301 MDLQCFLLSHSHQHDGMAYVKVDCSCEPQOLIMQNEEASDYDDLDSMDVVRP 360

Qy 380 DDDNSPSFIQIRSVAKGHFKTWHYIAAEEDWDYAPLVLPDDRYSKYSOYLNNGPQIG 439

```

QY 1520 YQDLFPPTETSGSPGHLDLVEGSLLOTEGAIKWNEANRCKVPFLRVATESSAKTPSK 1579
DB 1501 YQDLFPPTETSGSPGHLDLVEGSLLOTEGAIKWNEANRCKVPFLRVATESSAKTPSK 1560
QY 1580 LLDPLANDMNYGTQIPKEWKSQBSPEKTAFFKQDITLSNACESHAAINEGQNK 1639
DB 1561 LLDPLANDMNYGTQIPKEWKSQBSPEKTAFFKQDITLSNACESHAAINEGQNK 1620
QY 1640 BIEVWAKOGTERLCSQNPVLRKHOREITRTTLOSDBEIDYDDTISVEMKEDPDIY 1699
DB 1621 BIEVWAKOGTERLCSQNPVLRKHOREITRTTLOSDBEIDYDDTISVEMKEDPDIY 1680
QY 1700 DEDENQSPESFQKTRHVFIAAVERLDYGSSSPHVLNRQAQSVQPKKVVQBEFTD 1759
DB 1681 DEDENQSPESFQKTRHVFIAAVERLDYGSSSPHVLNRQAQSVQPKKVVQBEFTD 1740
QY 1760 GSFTQPLYRGELNEHLGLGYPYRAEVEDNIMVTPRNOASRPSPYSYSLISYEDQOQGA 1819
DB 1741 GSFTQPLYRGELNEHLGLGYPYRAEVEDNIMVTPRNOASRPSPYSYSLISYEDQOQGA 1800
QY 1820 EPRKNFVKPNETKTYFWKQVHHMPTKDBFCKAWAYFSDVDLEKDVHSGLIGPLLVCHT 1879
DB 1801 EPRKNFVKPNETKTYFWKQVHHMPTKDBFCKAWAYFSDVDLEKDVHSGLIGPLLVCHT 1860
QY 1880 NTLNPAHGRQVTVQEFALFFTIFDETKSWYPTENMERNCRAPCNIQMEDPTFKENYRPHA 1939
DB 1861 NTLNPAHGRQVTVQEFALFFTIFDETKSWYPTENMERNCRAPCNIQMEDPTFKENYRPHA 1920
QY 1940 INGYIMDTLPGLVMAQDORIRWYLLSMGSENENIHSIHFSCHVFTVRKKEEYKMALYNLYP 1999
DB 1921 INGYIMDTLPGLVMAQDORIRWYLLSMGSENENIHSIHFSCHVFTVRKKEEYKMALYNLYP 1980
QY 2000 GVFEVEMLPKAGIWRVECLIGEHLHAGMSTFLVYNSKCOPTPLGMASGHTRDFOITAS 2059
DB 1981 GVFEVEMLPKAGIWRVECLIGEHLHAGMSTFLVYNSKCOPTPLGMASGHTRDFOITAS 2040
QY 2060 GOYGOWAPKAPLAPLHYSGSINAWSTKEPFSWKVDLLAPMIHGIKTQGARQKFSLSYISQ 2119
DB 2041 GOYGOWAPKAPLAPLHYSGSINAWSTKEPFSWKVDLLAPMIHGIKTQGARQKFSLSYISQ 2100
QY 2120 FIIMYSLDGKKWQTYRGNSTGTLNMFVFGNVDSSGIXKHNIENPPIIARVIRLHPHTYSIRS 2179
DB 2101 FIIMYSLDGKKWQTYRGNSTGTLNMFVFGNVDSSGIXKHNIENPPIIARVIRLHPHTYSIRS 2160
QY 2180 TURBELMGCDLNSCNPGLGMSKASDAQITASSYFTNMFAWSPSKARLHLQGSNAWR 2239
DB 2161 TURBELMGCDLNSCNPGLGMSKASDAQITASSYFTNMFAWSPSKARLHLQGSNAWR 2220
QY 2240 POVNNPKEWLQVDFQKTMKVTGVTQGVKSLLTSMYKVEFLISSQDGHQWTLFPQNGKV 2299
DB 2221 POVNNPKEWLQVDFQKTMKVTGVTQGVKSLLTSMYKVEFLISSQDGHQWTLFPQNGKV 2280
QY 2300 KYFOCNOQSFTPVVNSLDPLILTRVIRIHPOSWHQIALRMEVLGCEAODLY 2351
DB 2281 KYFOCNOQSFTPVVNSLDPLILTRVIRIHPOSWHQIALRMEVLGCEAODLY 2332

Search completed: April 13, 2004, 14:08:32
Job time : 31.5 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.
OM protein - protein search, using sw model
Run on: April 13, 2004, 14:08:32 ; Search time 61.5 Seconds
(without alignments)
10051.498 Million cell updates/sec

Title: NP000123-328-355-581
Perfect score: 12415
Sequence: 1 MQIELSTCFCLLRFCFSA.....VWQIALRMEVLGCEAQDLY 2351

Scoring table: 3LOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1073127 seqs, 262937947 residues
Total number of hits satisfying chosen parameters: 1073127

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications AA:
1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep.*
10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep.*
11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/2/pubpaa/US09D_PUBCOMB.pep.*
13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep.*
14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep.*
15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/2/pubpaa/US10D_PUBCOMB.pep.*
17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES						
Result No.	Score	Query Match	Length	ID	Description	
1	12399	99.9	2351	14	US-10-132-829-4	Sequence 4, Appli
2	12399	99.9	2351	14	US-10-172-712-27	Sequence 27, Appli
3	12399	99.9	2351	14	US-10-133-907-4	Sequence 4, Appli
4	12399	99.8	2351	12	US-10-411-037-30	Sequence 30, Appli
5	12399	99.8	2351	12	US-10-411-026-30	Sequence 30, Appli
6	12295	99.0	2332	15	US-10-360-101-229	Sequence 229, App
7	12282	98.9	2332	9	US-09-957-641-2	Sequence 2, Appli
8	12282	98.9	2332	14	US-10-187-319-2	Sequence 2, Appli
9	12282	98.9	2332	14	US-10-131-510A-2	Sequence 2, Appli
10	12282	98.9	2332	15	US-10-445-235-2	Sequence 2, Appli
11	12282	98.9	2332	16	US-10-239-498A-2	Sequence 2, Appli
12	8820	71.0	2319	14	US-10-187-319-6	GENERAL INFORVA
13	8820	71.0	2319	14	US-10-131-510A-6	Sequence 6, Appli
14	8057	64.9	2133	14	US-10-187-319-37	Sequence 37, Appli
15	8057	64.9	2133	14	US-10-131-510A-37	Sequence 37, Appli

16	7393	59.5	1471	12	US-10-681-970-2	Sequence 2, Appli
17	7393	59.5	1471	13	US-10-095-718-2	Sequence 2, Appli
18	7294	58.8	1459	16	US-10-239-498A-13	Sequence 13, Appli
19	7284	58.7	1459	16	US-10-239-498A-15	Sequence 15, Appli
20	7281	58.6	1459	16	US-10-239-498A-4	Sequence 4, Appli
21	7208	58.1	1438	13	US-10-006-091-1	Sequence 1, Appli
22	7208	58.1	1438	13	US-10-047-257-1	Sequence 1, Appli
23	7208	58.1	1438	14	US-10-225-900-1	Sequence 1, Appli
24	6253	50.4	1431	12	US-10-681-970-4	Sequence 4, Appli
25	6253	50.4	1431	13	US-10-095-718-4	Sequence 4, Appli
26	6103	49.2	1443	14	US-10-187-319-39	Sequence 39, Appli
27	6103	49.2	1443	14	US-10-131-510A-39	Sequence 39, Appli
28	2825.5	22.7	2224	14	US-10-115-563-14	Sequence 14, Appli
29	2823.5	22.7	2196	15	US-10-360-101-259	Sequence 259, App
30	2823.5	22.7	2224	14	US-10-172-712-31	Sequence 31, Appli
31	1723	13.9	368	14	US-10-187-319-4	Sequence 4, Appli
32	1723	13.9	368	14	US-10-131-510A-4	Sequence 4, Appli
33	1340.5	10.8	1160	12	US-10-147-493-234	Sequence 234, App
34	1340.5	10.8	1160	12	US-10-145-127-234	Sequence 234, App
35	1340.5	10.8	1160	12	US-10-160-503-234	Sequence 234, App
36	1340.5	10.8	1160	12	US-10-143-118-234	Sequence 234, App
37	1340.5	10.8	1160	12	US-10-144-993-234	Sequence 234, App
38	1340.5	10.8	1160	12	US-10-158-787-234	Sequence 234, App
39	1340.5	10.8	1160	12	US-10-140-024-234	Sequence 234, App
40	1340.5	10.8	1160	14	US-10-028-072-234	Sequence 234, App
41	1340.5	10.8	1160	14	US-10-121-049-234	Sequence 234, App
42	1340.5	10.8	1160	14	US-10-123-904-234	Sequence 234, App
43	1340.5	10.8	1160	14	US-10-140-470-234	Sequence 234, App
44	1340.5	10.8	1160	14	US-10-175-748-234	Sequence 234, App
45	1340.5	10.8	1160	14	US-10-176-918-234	Sequence 234, App

ALIGNMENTS

RESULT 1
US-10-132-829-4
; Sequence 4, Application US/10132829
; Publication No. US20030044982A1
; GENERAL INFORMATION:
; APPLICANT: Chien, Kenneth R
; APPLICANT: Hoshijima, Masahiko
; TITLE OF INVENTION: Method to treat hemophilia by hepatic gene transfer of Factor VIII
; TITLE OF INVENTION: with vesicle vector
; FILE REFERENCE: 6627-PA1170
; CURRENT APPLICATION NUMBER: US/10/132,829
; CURRENT FILING DATE: 2002-04-25
; PRIOR FILING DATE: 2002-04-25
; PRIOR FILING DATE: 2001-04-25
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 4
; LENGTH: 2351
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-10-132-829-4

Query Match		99.9%	Score 12399;	DB 14;	Length 2351;
Best Local Similarity		99.9%	Pred. No. 0;		
Matches 2348;		Conservative 1;	Mismatches 2;	Indels 0;	Gaps 0;
Qy	1	MQIELSTCFCLLRFCFSA	TRRYLGA	VELSWD	YMQSLGELPVDARFPFRVKSPPFN 60
Db	1	MQIELSTCFCLLRFCFSA	TRRYLGA	VELSWD	YMQSLGELPVDARFPFRVKSPPFN 60
Qy	61	TSVYKKTLEVFETDHLPIAK	PRPFWMLL	GLGTIQAEVDTVVITLKN	MAFVSLHAV 120
Db	61	TSVYKKTLEVFETDHLPIAK	PRPFWMLL	GLGTIQAEVDTVVITLKN	MAFVSLHAV 120
Qy	121	GVSYWKASGAEYDDQTSORE	KEDDKVFP	GGSHYVWVLKENG	PMASDFLCITYSLH 180
Db	121	GVSYWKASGAEYDDQTSORE	KEDDKVFP	GGSHYVWVLKENG	PMASDFLCITYSLH 180

181 VDLVKDNLNSGLIGALLVCREGSLAKEKTQTLHKFIILLFAVFDGSKSESTKNSLMQDRD 240
181 VDLVKDNLNSGLIGALLVCREGSLAKEKTQTLHKFIILLFAVFDGSKSESTKNSLMQDRD 240
241 AASARAWPKGHTVNGVYNNRSLPGLIGCHRSVYVHVHVGMTTPVHGFPLEGHTFLVRNH 300
241 AASARAWPKGHTVNGVYNNRSLPGLIGCHRSVYVHVHVGMTTPVHGFPLEGHTFLVRNH 300
301 QOASLEISPIITELTAOTLMLDGLGOFILCHLSSHOHDGMEAYVAVDSCPPEPOLIMKNE 360
301 QOASLEISPIITELTAOTLMLDGLGOFILCHLSSHOHDGMEAYVAVDSCPPEPOLIMKNE 360
361 EAEDYDDDLTDSEMDVVRFDNNSPFTQIRSAVAKGPKTWVHVIAAEEDWDVAPLVL 420
361 EAEDYDDDLTDSEMDVVRFDNNSPFTQIRSAVAKGPKTWVHVIAAEEDWDVAPLVL 420
421 PDRSYKQVYLNNQFORIGRKYKVRMAYTDETFKTRAIQHSYGILGPLYGEVGDTL 480
421 PDRSYKQVYLNNQFORIGRKYKVRMAYTDETFKTRAIQHSYGILGPLYGEVGDTL 480
481 LIIFKNOASREYNILYPHGITDVRPLYSRRLPKGVKHLKDFILPCEIFPKYKWTVTVEDGP 540
481 LIIFKNOASREYNILYPHGITDVRPLYSRRLPKGVKHLKDFILPCEIFPKYKWTVTVEDGP 540
541 TKSDPRCLTRYSSFVNMERDLASGLIGPLLIICYKESVDQKNOIVSDKRNVLIFSVPDE 600
541 TKSDPRCLTRYSSFVNMERDLASGLIGPLLIICYKESVDQKNOIVSDKRNVLIFSVPDE 600
601 NRSYLTENIORFLPNPAGVQLEBPPEFOASNIMESINGVYVDSLSVCLHEVAYWYLS 660
601 NRSYLTENIORFLPNPAGVQLEBPPEFOASNIMESINGVYVDSLSVCLHEVAYWYLS 660
661 ICAQTDLSVFFSGVTFKHKVYEDTLTLFPFSGBTFVMSMENFGLWILGCHNSDFNRG 720
661 ICAQTDLSVFFSGVTFKHKVYEDTLTLFPFSGBTFVMSMENFGLWILGCHNSDFNRG 720
721 MTALLKVSSCDKGTDDYEDSYEDISAYLLSKNAIEPRFSQNSRRHSTPQKQFNATTI 780
721 MTALLKVSSCDKGTDDYEDSYEDISAYLLSKNAIEPRFSQNSRRHSTPQKQFNATTI 780
781 PENDIEKTDPAFAHTEPKTQNTSSSLLMLLQSPPHGLISLSDIOEAKYETFSDDPS 840
781 PENDIEKTDPAFAHTEPKTQNTSSSLLMLLQSPPHGLISLSDIOEAKYETFSDDPS 840
841 PGASDNNLSGEMTHFRPQLHSGDMVTPPSGLQLRLNEXLGTAAATELKKLPFKVSST 900
841 PGASDNNLSGEMTHFRPQLHSGDMVTPPSGLQLRLNEXLGTAAATELKKLPFKVSST 900
901 SNNLISTIPSDNLAAGTNTSLGPPSPMPHYDSQDITTLFGKKSSPLTESGGPLSLSEE 960
901 SNNLISTIPSDNLAAGTNTSLGPPSPMPHYDSQDITTLFGKKSSPLTESGGPLSLSEE 960
961 NNDKSLLESGLMNSOESSWGNVSTESGRLPKGRAGHPALLTKDNALFKVYSLLKTN 1020
961 NNDKSLLESGLMNSOESSWGNVSTESGRLPKGRAGHPALLTKDNALFKVYSLLKTN 1020
1021 KTSNNSATNRKTHIDCPILLSIENSFVQNILESDETFPKVTPPLIHRMLMDKNAATALRL 1080
1021 KTSNNSATNRKTHIDCPILLSIENSFVQNILESDETFPKVTPPLIHRMLMDKNAATALRL 1080
1081 NMSNKTITSSKNMVMVQKKEGPIPDQAQNPMSFPKMLFLPESARWQRTGKNSLNSG 1140
1081 NMSNKTITSSKNMVMVQKKEGPIPDQAQNPMSFPKMLFLPESARWQRTGKNSLNSG 1140
1141 QGPPSPKQLVSLGPEKSVBQGNFLSKNKVVVGKGBFTKDVGLKEMVFPSSRNLFITNLDN 1200
1141 QGPPSPKQLVSLGPEKSVBQGNFLSKNKVVVGKGBFTKDVGLKEMVFPSSRNLFITNLDN 1200
1201 LHEANTHNOEKKIQBSIEKKTLLIQENVVLPOIHTVTGTPKNFMKNLFLLLSTRQNVESYD 1260
1201 LHEANTHNOEKKIQBSIEKKTLLIQENVVLPOIHTVTGTPKNFMKNLFLLLSTRQNVESYD 1260
1261 GAYAPVLQDPRSLNDSNTKTKHTAHFSKGBEENLEGLGNQTKQIVEXYACTTRISPN 1320

1261 GAYAPVLQDPRSLNDSNTKTKHTAHFSKGBEENLEGLGNQTKQIVEXYACTTRISPN 1320
1321 SQQNFTQSRKALKQFRLPLETELEKRIIVDDTSTQNSKNMKHLTPSTLTQIDVNEKE 1380
1321 SQQNFTQSRKALKQFRLPLETELEKRIIVDDTSTQNSKNMKHLTPSTLTQIDVNEKE 1380
1381 KGAITQSPISDCITRSHSIPQANRSPLPIAKVSPFSPIRIYLRVLFQDNSSHLPAASY 1440
1381 KGAITQSPISDCITRSHSIPQANRSPLPIAKVSPFSPIRIYLRVLFQDNSSHLPAASY 1440
1441 RKXDSGVQESHEFLQAKAKNLSIALITLMTQDQREVSGLSGTSATNSVYTKVENVTLVP 1500
1441 RKXDSGVQESHEFLQAKAKNLSIALITLMTQDQREVSGLSGTSATNSVYTKVENVTLVP 1500
1501 KPDLPKTSGVAVELLPKVHIYQKDLFPETNSGSPGHLDLVEGSLLOQTEGAIKWNEANRP 1560
1501 KPDLPKTSGVAVELLPKVHIYQKDLFPETNSGSPGHLDLVEGSLLOQTEGAIKWNEANRP 1560
1561 GKVPFLRVATESSAKTPSKLLDPLANVNYGTQIPKEEWKSQEKSPKTAFFKXDTILSL 1620
1561 GKVPFLRVATESSAKTPSKLLDPLANVNYGTQIPKEEWKSQEKSPKTAFFKXDTILSL 1620
1621 NACESNHAIIAINEGQNKPEIEVTWAKQGTRELCSONPPVLKXHQREIIRTTLOSQEE 1680
1621 NACESNHAIIAINEGQNKPEIEVTWAKQGTRELCSONPPVLKXHQREIIRTTLOSQEE 1680
1681 IDYDDTISVEMKXEDFDIYDEDENQSPRSFOKTRHYFIAAVERLMDYGMSSSPHVLNR 1740
1681 IDYDDTISVEMKXEDFDIYDEDENQSPRSFOKTRHYFIAAVERLMDYGMSSSPHVLNR 1740
1741 AQSGSVPOFKKVPQFETDGSFTQPLYRGELNHLGLLGYIRAEVEDNINMTFRNOASR 1800
1741 AQSGSVPOFKKVPQFETDGSFTQPLYRGELNHLGLLGYIRAEVEDNINMTFRNOASR 1800
1801 PYSYSSILSYEBDORQASPRKNFYKNETKTYFWKVOQHMAPTKDEPCKAWAYSDV 1860
1801 PYSYSSILSYEBDORQASPRKNFYKNETKTYFWKVOQHMAPTKDEPCKAWAYSDV 1860
1861 DLEKDVHSGILGPLLVTCHTNTLPAHQRTVQVBOFALFFTFIDETKSWFYFENNERCRA 1920
1861 DLEKDVHSGILGPLLVTCHTNTLPAHQRTVQVBOFALFFTFIDETKSWFYFENNERCRA 1920
1921 PCNTQMBDPTFKENYRPFALINGYIMDTPLGLVMAQDQRIRWYLLSMGSENENIHSFSGH 1980
1921 PCNTQMBDPTFKENYRPFALINGYIMDTPLGLVMAQDQRIRWYLLSMGSENENIHSFSGH 1980
1981 VFTVRKKEEKWALYNLYPGVFETVEMLPKAGIWRVVECLIGELHAGMSTFLVYSNKC 2040
1981 VFTVRKKEEKWALYNLYPGVFETVEMLPKAGIWRVVECLIGELHAGMSTFLVYSNKC 2040
2041 QTPLGASGHIRDFQITASGOYGQWAPKLARLHYSGSINAWSTKEPFSWIKVDLLAPMII 2100
2041 QTPLGASGHIRDFQITASGOYGQWAPKLARLHYSGSINAWSTKEPFSWIKVDLLAPMII 2100
2101 HGIKTQARQXFSLLYSQFIIMYSLDCKKQVTRGNSGTGLMVPFGNVDSGSKHNIFN 2160
2101 HGIKTQARQXFSLLYSQFIIMYSLDCKKQVTRGNSGTGLMVPFGNVDSGSKHNIFN 2160
2161 PPIIARYRLPHPTHYSIRSTRLXELMGCDLNSCNPIMGESKAISSDAQITASSYFTNFA 2220
2161 PPIIARYRLPHPTHYSIRSTRLXELMGCDLNSCNPIMGESKAISSDAQITASSYFTNFA 2220
2221 TWSFSKARLHQGRSNARWPQVNNPKEWLQVDFQTKMKTGVTTCQVKSLTTSYVKEFL 2280
2221 TWSFSKARLHQGRSNARWPQVNNPKEWLQVDFQTKMKTGVTTCQVKSLTTSYVKEFL 2280
2281 ISSSQDGHQWTLFTQNGKVKVQGNQDSFTPVNSLDPPLLTRILRHPQSWHQAIALRM 2340
2281 ISSSQDGHQWTLFTQNGKVKVQGNQDSFTPVNSLDPPLLTRILRHPQSWHQAIALRM 2340
2341 EVLGCERQDLY 2351

Qy 1741 AQSQSVQFQKVVQFEFTDGSFTQPLRGELNEHLGLGPYIRAEVDEINMVFERNQASR 1800
Db 1741 AQSQSVQFQKVVQFEFTDGSFTQPLRGELNEHLGLGPYIRAEVDEINMVFERNQASR 1800
Qy 1801 PYSYSSLIISIEEDQOQAEPRKQFVKNETKTYFWKVQHMACTKDEFFDCKAWAYFSDV 1860
Db 1801 PYSYSSLIISIEEDQOQAEPRKQFVKNETKTYFWKVQHMACTKDEFFDCKAWAYFSDV 1860
Qy 1861 DLEKDVHSGLLGPIILVCHTNTILNPAHQROVTVQBFALPFTTFDFTKSYFTENNERCRA 1920
Db 1861 DLEKDVHSGLLGPIILVCHTNTILNPAHQROVTVQBFALPFTTFDFTKSYFTENNERCRA 1920
Qy 1921 PCNIQMEDPTFKENYRFHAINGYIMDTPLGLVMAQDQIRWYLLSMGSENENIHSIFSGH 1980
Db 1921 PCNIQMEDPTFKENYRFHAINGYIMDTPLGLVMAQDQIRWYLLSMGSENENIHSIFSGH 1980
Qy 1981 VFTVRKKEBYKVALYNLYPGVFEVEMLPKAGIRWVECLIGEHLHAGMSTLFLVYSNKC 2040
Db 1981 VFTVRKKEBYKVALYNLYPGVFEVEMLPKAGIRWVECLIGEHLHAGMSTLFLVYSNKC 2040
Qy 2041 QTPGLMASGHIRDFOITASGGYQGWAPKARLHYSGSINAWSTKEPFSWIKVDLLAPMII 2100
Db 2041 QTPGLMASGHIRDFOITASGGYQGWAPKARLHYSGSINAWSTKEPFSWIKVDLLAPMII 2100
Qy 2101 HGIKTQGARQFSSLYISQFIIMYSLDGKKQWTVRGNSTGTLMVPPGQVNDSSGKHNIEN 2160
Db 2101 HGIKTQGARQFSSLYISQFIIMYSLDGKKQWTVRGNSTGTLMVPPGQVNDSSGKHNIEN 2160
Qy 2161 PPIIARYIRLHETHYSIPSTLMEMLGCDLNSCPLGWSKASDAOITASSYFTNMFA 2220
Db 2161 PPIIARYIRLHETHYSIPSTLMEMLGCDLNSCPLGWSKASDAOITASSYFTNMFA 2220
Qy 2221 TWSPSKARLHQGRSNAWRPQVNNPKEMQLQVDFQKTMKVTVGTTQGVKSLLTSMYVKEFL 2280
Db 2221 TWSPSKARLHQGRSNAWRPQVNNPKEMQLQVDFQKTMKVTVGTTQGVKSLLTSMYVKEFL 2280
Qy 2281 ISSQDGHQWTLFPONGKVKVFGQNDQSTPVPVNSLDPLITRYLRHHPQSWVHQIALRM 2340
Db 2281 ISSQDGHQWTLFPONGKVKVFGQNDQSTPVPVNSLDPLITRYLRHHPQSWVHQIALRM 2340
Qy 2341 EVLGCFAODLY 2351
Db 2341 EVLGCFAODLY 2351

RESULT 3

US-10-133-907-4
; Sequence 4, Application US/10133907
; Publication No. US20030195223A1
; GENERAL INFORMATION:
; APPLICANT: Chien, Kenneth R
; APPLICANT: Hoshijima, Masahiko
; TITLE OF INVENTION: Method to treat hemophilia by hepatic gene transfer of Factor VIII
; TITLE OF INVENTION: Method to treat hemophilia by hepatic gene transfer of Factor VIII
; TITLE OF INVENTION: Method to treat hemophilia by hepatic gene transfer of Factor VIII
; FILE REFERENCE: 6627-Pall170
; CURRENT APPLICATION NUMBER: US/10/133,907
; CURRENT FILING DATE: 2002-04-25
; PRIOR FILING DATE: 2002-04-25
; PRIOR FILING DATE: 2001-04-25
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 4
; LENGTH: 2351
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-133-907-4

Query Match 99.5%; Score 12399; DB 14; Length 2351;
Best Local Similarity 99.5%; Pred. No. 0;
Matches 2348; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 MQIELSTCFCLLRFCFSATRRYYLGAVELSWDYMOSDLGELPVDARFPRVPKSPFN 60

Db 1 MQIELSTCFCLLRFCFSATRRYYLGAVELSWDYMOSDLGELPVDARFPRVPKSPFN 60
Qy 61 TSVVYKXKTLVEFTDHLFNIAKRRPPMMGLLGPITIOAEVYDVTVVITLKNWASHPVSLHAV 120
Db 61 TSVVYKXKTLVEFTDHLFNIAKRRPPMMGLLGPITIOAEVYDVTVVITLKNWASHPVSLHAV 120
Qy 121 GVSYWKASEGAEYDDQTSOREKDDKVFPQGSSTYVQVLKENGPMASDPLCLTYSYLSH 180
Db 121 GVSYWKASEGAEYDDQTSOREKDDKVFPQGSSTYVQVLKENGPMASDPLCLTYSYLSH 180
Qy 181 VDLVKDNLNSGLIGALVCRREGSLAKETOTLHKFILLFAVDEGKSWHSTKSLMOORD 240
Db 181 VDLVKDNLNSGLIGALVCRREGSLAKETOTLHKFILLFAVDEGKSWHSTKSLMOORD 240
Qy 241 AASARAMPKMHTVNGYVNRSLPGLIGCHRKSVYVHVIMGTTPVHGFIFLEHGTFLVRNH 300
Db 241 AASARAMPKMHTVNGYVNRSLPGLIGCHRKSVYVHVIMGTTPVHGFIFLEHGTFLVRNH 300
Qy 301 QOASLEISPTIFLTAQTLMLDLOGLLSCHISSHOHGMAYVYVDSCPPEPQIMQONE 360
Db 301 QOASLEISPTIFLTAQTLMLDLOGLLSCHISSHOHGMAYVYVDSCPPEPQIMQONE 360
Qy 361 EADYDDDLTDSEMDVVRFDKNSPSFIQIRSAKHPKTHVHIAAEEEDWDVAPLVLA 420
Db 361 EADYDDDLTDSEMDVVRFDKNSPSFIQIRSAKHPKTHVHIAAEEEDWDVAPLVLA 420
Qy 421 PDDRSYKSQVNLNNGPQIRGKVKVRFMAYTDEFTKREAIQHESGILGLLYGEVGDTL 480
Db 421 PDDRSYKSQVNLNNGPQIRGKVKVRFMAYTDEFTKREAIQHESGILGLLYGEVGDTL 480
Qy 481 LIIFKNOASRPYNTYPGIGITDVRPLYGRRLPKGVKHLKDPILPGEIFPKYKWTYVDEGP 540
Db 481 LIIFKNOASRPYNTYPGIGITDVRPLYGRRLPKGVKHLKDPILPGEIFPKYKWTYVDEGP 540
Qy 541 TKSDPRLCTRYYSFVNMERDLASGLIGPLLI CYKESVDQGNQIMSDKXNVILFSVFDE 600
Db 541 TKSDPRLCTRYYSFVNMERDLASGLIGPLLI CYKESVDQGNQIMSDKXNVILFSVFDE 600
Qy 601 NRSWYLTENIQRFLPNPAGVQLEDPEFQASNMHSINGYVDFSLQSLCYLHEVAYWILS 660
Db 601 NRSWYLTENIQRFLPNPAGVQLEDPEFQASNMHSINGYVDFSLQSLCYLHEVAYWILS 660
Qy 661 IGAQTDLSLVSPSGYTPKHVMYEDTLTLPFSETVPMSENENPCLWILGCHNSDFNRNG 720
Db 661 IGAQTDLSLVSPSGYTPKHVMYEDTLTLPFSETVPMSENENPCLWILGCHNSDFNRNG 720
Qy 721 MTALLKVSSCDKNTGDYVEDSYEDI SAYLLSKNNAIEPRSPQNSRHPSTTRQKQFNATTI 780
Db 721 MTALLKVSSCDKNTGDYVEDSYEDI SAYLLSKNNAIEPRSPQNSRHPSTTRQKQFNATTI 780
Qy 781 PENDIEKTPWFARHTPMFKIIONVSSDMLMLLQSPTPHGLSLSDLOEAKYETFSDDPS 840
Db 781 PENDIEKTPWFARHTPMFKIIONVSSDMLMLLQSPTPHGLSLSDLOEAKYETFSDDPS 840
Qy 841 PGADSNNSLSEMTHERPOLHSGDMVTPESGLOLRNEKLGTTAATLAKLDFKVSST 900
Db 841 PGADSNNSLSEMTHERPOLHSGDMVTPESGLOLRNEKLGTTAATLAKLDFKVSST 900
Qy 901 SNNLITSTIPDNLAAGTDNTSSLGPSPMPVHYDSQDQDTTLTFLGKSSPLTESGGFLSLEE 960
Db 901 SNNLITSTIPDNLAAGTDNTSSLGPSPMPVHYDSQDQDTTLTFLGKSSPLTESGGFLSLEE 960
Qy 961 NDSKLLBSGLMNSQESSWGKNVSTESGRFLFKRAHGPALLTKDNALFKVSI SLKTN 1020
Db 961 NDSKLLBSGLMNSQESSWGKNVSTESGRFLFKRAHGPALLTKDNALFKVSI SLKTN 1020
Qy 1021 KTSNNSATNRKTHIDGPSLLIENSPTWQNTLESDFKVTPLIHORMMDKNATLRL 1080
Db 1021 KTSNNSATNRKTHIDGPSLLIENSPTWQNTLESDFKVTPLIHORMMDKNATLRL 1080
Qy 1081 NMSNKTTSKKNMENVQOKEGPIZPPAQNPDM5FFKMLFLPESARMTQRTGKNSLNSG 1140
Db 1081 NMSNKTTSKKNMENVQOKEGPIZPPAQNPDM5FFKMLFLPESARMTQRTGKNSLNSG 1140

```
QY 1141 QGSPKQVSLGPEKSEGVNQFVSEKNNVVGKGEFTKDVGLKEMVPPSSRNFLTNLDN 1200
Db 1141 QGSPKQVSLGPEKSEGVNQFVSEKNNVVGKGEFTKDVGLKEMVPPSSRNFLTNLDN 1200
QY 1201 LHENNTNQKKEQEBIEKKEKTELQENWVLPQIHVTGTGNFMKNLFLLLSTRQNVGSDYD 1260
Db 1201 LHENNTNQKKEQEBIEKKEKTELQENWVLPQIHVTGTGNFMKNLFLLLSTRQNVGSDYD 1260
QY 1261 GAYAPVLODFRS.LNDSNTRKKTAAHPSKKGEPEENLEGLGNQTKQIVKCYACTRIISPNT 1320
Db 1261 GAYAPVLODFRS.LNDSNTRKKTAAHPSKKGEPEENLEGLGNQTKQIVKCYACTRIISPNT 1320
QY 1321 SQQNFVTRSKRALQKRLPLEETEELKRIIVDDTSTQWSKNMKHLTPSTLTQIDYNEKE 1380
Db 1321 SQQNFVTRSKRALQKRLPLEETEELKRIIVDDTSTQWSKNMKHLTPSTLTQIDYNEKE 1380
QY 1381 KGALTQSP.LSCLUTRSHSIPOANRSPPIAKVSSFPSPRIPIYLRVL.FQNSSHLPAASY 1440
Db 1381 KGALTQSP.LSCLUTRSHSIPOANRSPPIAKVSSFPSPRIPIYLRVL.FQNSSHLPAASY 1440
QY 1441 RKXDSGVQESASHFLQGAKNLSLAITLLEMTGDOREVSLGTSATNSVYKVKVENTVLP 1500
Db 1441 RKXDSGVQESASHFLQGAKNLSLAITLLEMTGDOREVSLGTSATNSVYKVKVENTVLP 1500
QY 1501 KPDLPKTSGKVLLPKVHIYQKDLFPPTETSNSSPGHLDLVEGSLLOGTGEGAIKWNEANRP 1560
Db 1501 KPDLPKTSGKVLLPKVHIYQKDLFPPTETSNSSPGHLDLVEGSLLOGTGEGAIKWNEANRP 1560
QY 1561 GKVPFLRVATESSAKTSPKLLDPLAWNHYGTQPKKEWKSQKSPKTAFAKKXDTILSL 1620
Db 1561 GKVPFLRVATESSAKTSPKLLDPLAWNHYGTQPKKEWKSQKSPKTAFAKKXDTILSL 1620
QY 1621 NACSNHAIAINNGQNKPELEVWAKOGRTERLCSQNPVVLKXHQREIITRTILQSQEE 1680
Db 1621 NACSNHAIAINNGQNKPELEVWAKOGRTERLCSQNPVVLKXHQREIITRTILQSQEE 1680
QY 1681 IDYDDTISVEMKKEDFDIYDEDENQSPRSFOKTRHYFIAAVERLWDYGMSSSPHVLNR 1740
Db 1681 IDYDDTISVEMKKEDFDIYDEDENQSPRSFOKTRHYFIAAVERLWDYGMSSSPHVLNR 1740
QY 1741 AQSQSVQFKVQVQEFDDGFTQPLRGELNEHLGLLGPYIRAEVDENIMVYTRNQASR 1800
Db 1741 AQSQSVQFKVQVQEFDDGFTQPLRGELNEHLGLLGPYIRAEVDENIMVYTRNQASR 1800
QY 1801 PYSFYSLISYEDORQAEPRKPVKNETKTYFKVQHMAPTKDEDFCKAWAYFSDV 1860
Db 1801 PYSFYSLISYEDORQAEPRKPVKNETKTYFKVQHMAPTKDEDFCKAWAYFSDV 1860
QY 1861 DLEKDVHSLIGPLLVCHTNTLNPAGHQVTVQEFALFTTIFDETQSWYFTENMERNCR 1920
Db 1861 DLEKDVHSLIGPLLVCHTNTLNPAGHQVTVQEFALFTTIFDETQSWYFTENMERNCR 1920
QY 1921 PCNTOMEDPTPKENVRPHAINGYIMDTLPGVMAQDQIRWYLLSMGSENHISHPGSH 1980
Db 1921 PCNTOMEDPTPKENVRPHAINGYIMDTLPGVMAQDQIRWYLLSMGSENHISHPGSH 1980
QY 1981 VFTVRKKEEYKVALYNLYPGVFETVEMLPKSGAGIWRVECLICEHLHAGMSTLFLVYSNKC 2040
Db 1981 VFTVRKKEEYKVALYNLYPGVFETVEMLPKSGAGIWRVECLICEHLHAGMSTLFLVYSNKC 2040
QY 2041 QTPCLNAGSHTRDFOITASGOYQWAPKARLHYSGS.NAMSTKBPFPFWIKVDLLAPMII 2100
Db 2041 QTPCLNAGSHTRDFOITASGOYQWAPKARLHYSGS.NAMSTKBPFPFWIKVDLLAPMII 2100
QY 2101 HGKITQGAOKFESSLYISOFIIMYSIDGKKWOTYRGNSTGLMWFFGNDVSSGIXHNIEN 2160
Db 2101 HGKITQGAOKFESSLYISOFIIMYSIDGKKWOTYRGNSTGLMWFFGNDVSSGIXHNIEN 2160
QY 2161 PPIIARYIRLHPTHYSIRSLRMELMGCDLNSCSMPLGMESKAISDAQITASSYFTNMFA 2220
Db 2161 PPIIARYIRLHPTHYSIRSLRMELMGCDLNSCSMPLGMESKAISDAQITASSYFTNMFA 2220
```

RESULT 4

```
US-10-411-037-30
; Sequence 30, Application US/10411037
; Publication No. US20040043446A1
; GENERAL INFORMATION:
; APPLICANT: Neose Technologies, Inc.
; APPLICANT: Defress, Shawn
; APPLICANT: Zopf, David
; APPLICANT: Bayer, Robert
; APPLICANT: Hakes, David
; APPLICANT: Chen, Xi
; APPLICANT: Rowe, Caryn
; TITLE OF INVENTION: ALPHA GALACTOSIDASE A: REMODELING AND GLYCOCONJUGATION OF ALPHA
; TITLE OF INVENTION: GALACTOSIDASE A
; FILE REFERENCE: 040853-01-5082
; CURRENT APPLICATION NUMBER: US/10/411,037
; CURRENT FILING DATE: 2003-04-09
; PRIOR APPLICATION NUMBER: US 60/328,523
; PRIOR FILING DATE: 2001-10-10
; PRIOR APPLICATION NUMBER: US 60/344,692
; PRIOR FILING DATE: 2001-10-29
; PRIOR APPLICATION NUMBER: US 60/387,292
; PRIOR FILING DATE: 2002-06-07
; PRIOR APPLICATION NUMBER: US 60/391,777
; PRIOR FILING DATE: 2002-06-25
; PRIOR APPLICATION NUMBER: US 60/396,594
; PRIOR FILING DATE: 2002-07-17
; PRIOR APPLICATION NUMBER: US 60/404,249
; PRIOR FILING DATE: 2002-08-16
; PRIOR APPLICATION NUMBER: US 60/407,527
; PRIOR FILING DATE: 2002-08-28
; NUMBER OF SEQ ID NOS: 75
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 30
; LENGTH: 2351
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-411-037-30
```

Query Match 99.8%; Score 12390; DB 12; Length 2351;

Best Local Similarity 99.8%; Pseq. No. 0;

Matches 2347; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 MQIELSTCFPLCLARFCFSATRRYYLGAVELSWDYMOSDLGELPVDARFPFRVPSKSPFN 60

Db 1 MQIELSTCFPLCLARFCFSATRRYYLGAVELSWDYMOSDLGELPVDARFPFRVPSKSPFN 60

QY 61 TSVVYKKTILVEFTDHLFNIAKPPPKWGLLGPTIQAEVYDVTITLKNMASHPVSLHAV 120

Db 61 TSVVYKKTILVEFTDHLFNIAKPPPKWGLLGPTIQAEVYDVTITLKNMASHPVSLHAV 120

QY 121 GVSYWKASEGAEYDDQTSQREKDDKVPFGSSHYYVQVILKENGPMASDPLCLTYSYLSH 180

Db 121 GVSYWKASEGAEYDDQTSQREKDDKVPFGSSHYYVQVILKENGPMASDPLCLTYSYLSH 180

QY 181 VDLVKDNLNSGLIGALLVCREGSLAKEKTQTLHKFILLFAVEDEGKSWHSETKNSLMQORD 240

Db 181 VDLVKDNLNSGLIGALLVCREGSLAKEKTQTLHKFILLFAVEDEGKSWHSETKNSLMQORD 240

QY 241 AASRAWPKKETHVANGYVNRSLPGLIGCHRKSVYWHVIGMTTPEVHSIFLEGHTFLVRNH 300

Db 241 AASARANKMETUNGYNRSIPLGLIGHCKRKYVWJICMGITPEVHSIFLBSHTFLVRNH 300
Qy 301 ROASLEISPIITFLTAQILLMDLGOFLLSCHSSHOHDMGMEYKVVDSCEPBPOLIMKNE 360
Db 301 ROASLEISPIITFLTAQILLMDLGOFLFCHSSHOHDMGMEYKVVDSCEPBPOLIMKNE 360
Qy 361 BAEDYDDDLTDEMDVVRFDONSPSFIORSVAKKPKTWVHYIAABEEDWDVAPLVLA 420
Db 361 BAEDYDDDLTDEMDVVRFDONSPSFIORSVAKKPKTWVHYIAABEEDWDVAPLVLA 420
Qy 421 PDCRSYKQYLNGPQRIGRKYKVRFWAYTDETFKTREAIQHSBGILGLPYGEVGDTL 480
Db 421 PDCRSYKQYLNGPQRIGRKYKVRFWAYTDETFKTREAIQHSBGILGLPYGEVGDTL 480
Qy 481 LIIFPNQASRPYNYPHGITDVRPLYSRPLPKGVKHLKDFILPCEIFPKYKWTVTEDGP 540
Db 481 LIIFPNQASRPYNYPHGITDVRPLYSRPLPKGVKHLKDFILPCEIFPKYKWTVTEDGP 540
Qy 541 TKSDPRCLTRYSSFNMERDLASGLIGPLIICYKESVDQKGNQIMSDKRNWILFSVFDE 600
Db 541 TKSDPRCLTRYSSFNMERDLASGLIGPLIICYKESVDQKGNQIMSDKRNWILFSVFDE 600
Qy 601 NRSWYLTENIQRFPNPKAGVQLEDEPFOASNMHSINGYVDSLSQLSVCLHEVAYWILS 660
Db 601 NRSWYLTENIQRFPNPKAGVQLEDEPFOASNMHSINGYVDSLSQLSVCLHEVAYWILS 660
Qy 661 IGAOTDFLUSVPSGVYTFKHKWVYEDTLTLPSPGETVFMSENPCGLWILGCHNSDFRNG 720
Db 661 IGAOTDFLUSVPSGVYTFKHKWVYEDTLTLPSPGETVFMSENPCGLWILGCHNSDFRNG 720
Qy 721 MTALLKVSCKDNTGYYEDSDYEDISAYLLSKNNAIEPRSPQNSRHPSTKQKFNATTI 780
Db 721 MTALLKVSCKDNTGYYEDSDYEDISAYLLSKNNAIEPRSPQNSRHPSTKQKFNATTI 780
Qy 781 PENDIEKTDPAHRTMPKTONVSSDLMLLRQSPTPHGLSLSDLOEAKYETPSDDPS 840
Db 781 PENDIEKTDPAHRTMPKTONVSSDLMLLRQSPTPHGLSLSDLOEAKYETPSDDPS 840
Qy 841 PGALDSNNLSGEMTHFRPOLHSGDMVTPSGLOLRNEKLGTAATLKLKLPKVSST 900
Db 841 PGALDSNNLSGEMTHFRPOLHSGDMVTPSGLOLRNEKLGTAATLKLKLPKVSST 900
Qy 901 SNNLTSTIPSONLAAGTDNTSLGPPSPVHYDSQDNTTLFGKKSPTESGGLSLSEE 960
Db 901 SNNLTSTIPSONLAAGTDNTSLGPPSPVHYDSQDNTTLFGKKSPTESGGLSLSEE 960
Qy 961 NNDKSLLESGLMNSOESSGKNVSTESGRLPKGAHGPALLTKDNALFKVSISSLKTN 1020
Db 961 NNDKSLLESGLMNSOESSGKNVSTESGRLPKGAHGPALLTKDNALFKVSISSLKTN 1020
Qy 1021 KTSNNSATNRKTHIDGPELLTENSPPVQNILESDETEPKVTPLTHDRMLMDKXATLRL 1080
Db 1021 KTSNNSATNRKTHIDGPELLTENSPPVQNILESDETEPKVTPLTHDRMLMDKXATLRL 1080
Qy 1081 NHMSNKTSSKNMVMQKKEGPIPPDAQNPDMSPFKMLFIPESARWIQRTGHKNSLNSG 1140
Db 1081 NHMSNKTSSKNMVMQKKEGPIPPDAQNPDMSPFKMLFIPESARWIQRTGHKNSLNSG 1140
Qy 1141 CGPSPKQLVSLGPEKSVGQNFLESBNKVVVGKGEPTKDVGLKEMVFPSSRNLPJTNLDN 1200
Db 1141 CGPSPKQLVSLGPEKSVGQNFLESBNKVVVGKGEPTKDVGLKEMVFPSSRNLPJTNLDN 1200
Qy 1201 LHENNTNHOEKKIQBEIEKKTLLIQENVVLPOIHTVTGPKNFMKNLFLLSRQNVGSGYD 1260
Db 1201 LHENNTNHOEKKIQBEIEKKTLLIQENVVLPOIHTVTGPKNFMKNLFLLSRQNVGSGYD 1260
Qy 1261 GAYAPVLQDFRSLNDSTNRTKXHTAHFSKKGBEENLEGJNQTKQIVEKYACTTRISPT 1320
Db 1261 GAYAPVLQDFRSLNDSTNRTKXHTAHFSKKGBEENLEGJNQTKQIVEKYACTTRISPT 1320
Qy 1321 SQONFVQSRKALKQFRPLPBEETELEKRIIVDDTSTQWSKNMKHLTPSTLTQIDYNEKE 1380

RESULT 5

US-10-411-026-30

Db 1321 SQONFVQSRKALKQFRPLPBEETELEKRIIVDDTSTQWSKNMKHLTPSTLTQIDYNEKE 1380
Qy 1381 KGALTQSPLDCLTRSHSIFQANRSPUPIAKVSPSPSIRPIYLRVLQONSSHLPAASY 1440
Db 1381 KGALTQSPLDCLTRSHSIFQANRSPUPIAKVSPSPSIRPIYLRVLQONSSHLPAASY 1440
Qy 1441 RKXDSGVQESSHFIQGAKKNNLSLAITLBMTDQREVSGSLGTSATNSVTYKKVENTVLP 1500
Db 1441 RKXDSGVQESSHFIQGAKKNNLSLAITLBMTDQREVSGSLGTSATNSVTYKKVENTVLP 1500
Qy 1501 KPDLPKTSGVVELLPKVHIYQKOLPPTETSGSPGHLDLVEGSLLOCTEGAIKWEANRP 1560
Db 1501 KPDLPKTSGVVELLPKVHIYQKOLPPTETSGSPGHLDLVEGSLLOCTEGAIKWEANRP 1560
Qy 1561 GKVPFLRVATESAKTSPKLLDPLAWNHYGTQIPKEWKSQEKSPKTAFFKKDITLSL 1620
Db 1561 GKVPFLRVATESAKTSPKLLDPLAWNHYGTQIPKEWKSQEKSPKTAFFKKDITLSL 1620
Qy 1621 NACESNHAIALNBNQKPEIEVTWAKQGRTERLCSQNPVPLKHHQREITRTTLOSQDEE 1680
Db 1621 NACESNHAIALNBNQKPEIEVTWAKQGRTERLCSQNPVPLKHHQREITRTTLOSQDEE 1680
Qy 1681 IDYDDTISVENKKEDDFDYDEDENQSPRSFQKTRHYFIAAVERPLMDYGYSSSHVLNR 1740
Db 1681 IDYDDTISVENKKEDDFDYDEDENQSPRSFQKTRHYFIAAVERPLMDYGYSSSHVLNR 1740
Qy 1741 AOSGSVPQFKVVFQEBTQDGSFTQPLYRGELNEHLGLLGPYIRAEVDNIMVTRNQASR 1800
Db 1741 AOSGSVPQFKVVFQEBTQDGSFTQPLYRGELNEHLGLLGPYIRAEVDNIMVTRNQASR 1800
Qy 1801 PYSYSSLISSVEEDQROGAEPRKNPVKNETKTVFVKVQHMAFTKQBFDCAKWAYFSDV 1860
Db 1801 PYSYSSLISSVEEDQROGAEPRKNPVKNETKTVFVKVQHMAFTKQBFDCAKWAYFSDV 1860
Qy 1861 DLEKDVHSGILGPLLVCHTNTLNPAHQSOVTVQBFALFFTFIDETKSWYFTENNERCRA 1920
Db 1861 DLEKDVHSGILGPLLVCHTNTLNPAHQSOVTVQBFALFFTFIDETKSWYFTENNERCRA 1920
Qy 1921 PCNIQMEDPTFKENYRPHAINGYINDTLPGLVMAQDQIRWYLLSMGNSNENIHSHFSGH 1980
Db 1921 PCNIQMEDPTFKENYRPHAINGYINDTLPGLVMAQDQIRWYLLSMGNSNENIHSHFSGH 1980
Qy 1981 VFTVRKKEBYKMALYNLYPGVFTVEMLPKAGIWRVECLIGEHLHAGMSTLFLVYSNKC 2040
Db 1981 VFTVRKKEBYKMALYNLYPGVFTVEMLPKAGIWRVECLIGEHLHAGMSTLFLVYSNKC 2040
Qy 2041 QTPLGMAAGHTRDFOITASGOYGOWAPKLAFLHYSGSINAMSTKEPFSWIKVDLLAPMI 2100
Db 2041 QTPLGMAAGHTRDFOITASGOYGOWAPKLAFLHYSGSINAMSTKEPFSWIKVDLLAPMI 2100
Qy 2101 HGIKTQARQAFSSLYISQFIIMYSLDGKKMQTYRGNSTGTLMVFFGNVDSGKHNIFN 2160
Db 2101 HGIKTQARQAFSSLYISQFIIMYSLDGKKMQTYRGNSTGTLMVFFGNVDSGKHNIFN 2160
Qy 2161 PPIIARVIRLPHTHYSIRSTRMELMGCDLNSCMPLGMEKASISDAQITASSYFTNMPA 2220
Db 2161 PPIIARVIRLPHTHYSIRSTRMELMGCDLNSCMPLGMEKASISDAQITASSYFTNMPA 2220
Qy 2221 TWSPSKARLHLQGSNAWRPOVNNPKWLQVDFOKTKVTVGTQGVKSLTSMYVKEFL 2280
Db 2221 TWSPSKARLHLQGSNAWRPOVNNPKWLQVDFOKTKVTVGTQGVKSLTSMYVKEFL 2280
Qy 2281 ISSQDGHQWTLFTQNGKVKVQGNQDSFTPVVNSLDPPLLTRYLRIHPQSWVHQIALRM 2340
Db 2281 ISSQDGHQWTLFTQNGKVKVQGNQDSFTPVVNSLDPPLLTRYLRIHPQSWVHQIALRM 2340
Qy 2341 EVLGCBAQDLY 2351
Db 2341 EVLGCBAQDLY 2351

/ Sequence 30, Application US/10411026
/ Publication No. US20040063911A1
/ GENERAL INFORMATION:
/ APPLICANT: Neose Technologies, Inc.
/ APPLICANT: DeFrees, Shawn
/ APPLICANT: Zopf, David
/ APPLICANT: Bayer, Robert
/ APPLICANT: Hakes, David
/ APPLICANT: Chen, Xi
/ TITLE OF INVENTION: PROTEIN REMODELING METHODS AND PROTEINS/PEPTIDES PRODUCED BY THE
/ TITLE OF INVENTION: METHODS
/ FILE REFERENCE: G40853-01-5053
/ CURRENT APPLICATION NUMBER: US/10/411,026
/ CURRENT FILING DATE: 2003-04-09
/ PRIOR APPLICATION NUMBER: US 60/328,523
/ PRIOR FILING DATE: 2001-10-10
/ PRIOR APPLICATION NUMBER: US 60/344,692
/ PRIOR FILING DATE: 2001-10-19
/ PRIOR APPLICATION NUMBER: US 60/387,292
/ PRIOR FILING DATE: 2002-06-07
/ PRIOR APPLICATION NUMBER: US 60/391,777
/ PRIOR FILING DATE: 2002-06-25
/ PRIOR APPLICATION NUMBER: US 60/396,594
/ PRIOR FILING DATE: 2002-07-17
/ PRIOR APPLICATION NUMBER: US 60/404,249
/ PRIOR FILING DATE: 2002-08-16
/ PRIOR APPLICATION NUMBER: US 60/407,527
/ PRIOR FILING DATE: 2002-08-28
/ NUMBER OF SEQ ID NOS: 75
/ SOFTWARE: PatentIn version 3.2
/ SEQ ID NO 30
/ LENGTH: 2351
/ TYPE: prt
/ ORGANISM: Homo sapiens
US-10-411-026-30

Query Match 99.8%; Score 12390; DB 12; Length 2351;
Best Local Similarity 99.8%; Pred.No. 0;
Matches 2347; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY	1	MOELSTCFELCLLRFCSATRRYLGAVLSWDYMQSDLGELPVDARFPRPKSPFN	60
DB	1	MOELSTCFELCLLRFCSATRRYLGAVLSWDYMQSDLGELPVDARFPRPKSPFN	60
QY	61	TSVYKTKLVEFTDHLFNIAKPRPPMGLLGPTIOAEVDTVVITLKNMASHPVSLHAV	120
DB	61	TSVYKTKLVEFTDHLFNIAKPRPPMGLLGPTIOAEVDTVVITLKNMASHPVSLHAV	120
QY	121	GVSYWKASEGAEDDQTSQREKEDKVPFGSHTYVWQVLKENGPMASDPLCLITYSLSH	180
DB	121	GVSYWKASEGAEDDQTSQREKEDKVPFGSHTYVWQVLKENGPMASDPLCLITYSLSH	180
QY	181	VDLVKOLNSGLIGALLVCRGSLAKETQTLHKFILLFAVFDGKSWHSETKNSLMODRD	240
DB	181	VDLVKOLNSGLIGALLVCRGSLAKETQTLHKFILLFAVFDGKSWHSETKNSLMODRD	240
QY	241	AAGARAWPKMHTVNGYVNRSLPGLIGCHRXSVYWHVICMGTTPEVHSIFLEGHTFLVRNH	300
DB	241	AAGARAWPKMHTVNGYVNRSLPGLIGCHRXSVYWHVICMGTTPEVHSIFLEGHTFLVRNH	300
QY	301	QASLEISPIITELTAQTLMDLQGLFLLSCHSHQHDGMEAYKVVDSCPEEPQLRMKNE	360
DB	301	QASLEISPIITELTAQTLMDLQGLFLLSCHSHQHDGMEAYKVVDSCPEEPQLRMKNE	360
QY	361	EADYDDDLTDSMDVVRFDNDSPSTQIRSVAKKPKTWVHVIAAEEEDWDYAPLVLA	420
DB	361	EADYDDDLTDSMDVVRFDNDSPSTQIRSVAKKPKTWVHVIAAEEEDWDYAPLVLA	420
QY	421	PDRSYKSOYVINGPQIRGRYKVRPMAYTDEFTKTRAIQHESGILGLLYGEVGDTL	480
DB	421	PDRSYKSOYVINGPQIRGRYKVRPMAYTDEFTKTRAIQHESGILGLLYGEVGDTL	480
QY	481	LIIFKQASRPYNIYPHGITDVRPLYSRRLPKGVKHLKDPFILPGEIFKYKWTVTVEDGP	540

DB	481	LIIFKQASRPYNIYPHGITDVRPLYSRRLPKGVKHLKDPFILPGEIFKYKWTVTVEDGP	540
QY	541	TKSDPECLTRYSSFVNVERDLASGLIGLLICYKESVDQGNQIMSDAKENVILFSVFDE	600
DB	541	TKSDPECLTRYSSFVNVERDLASGLIGLLICYKESVDQGNQIMSDAKENVILFSVFDE	600
QY	601	NRSWYLTENIORFLPNPAGVQLEDPEFQASNMHSINGVYFDSQLQSVCLHEVAYWYLS	660
DB	601	NRSWYLTENIORFLPNPAGVQLEDPEFQASNMHSINGVYFDSQLQSVCLHEVAYWYLS	660
QY	661	IGAQTDFLSVFFSGYTFKHVMYVEDTLTLPFSGGTVPFMSMENPGLWILGCHNSDPRNRG	720
DB	661	IGAQTDFLSVFFSGYTFKHVMYVEDTLTLPFSGGTVPFMSMENPGLWILGCHNSDPRNRG	720
QY	721	MTALLKVSSCDKNTGYYEDSVEDIISAYLLSKNATEPRFSFNSRHPSTROKQFNATTI	780
DB	721	MTALLKVSSCDKNTGYYEDSVEDIISAYLLSKNATEPRFSFNSRHPSTROKQFNATTI	780
QY	781	PENDIEKTDPFWAHRTMPKIQNVSSDLMMLLRQSPTEPHGLSLSDLOBAKYETFSDDPS	840
DB	781	PENDIEKTDPFWAHRTMPKIQNVSSDLMMLLRQSPTEPHGLSLSDLOBAKYETFSDDPS	840
QY	841	PGAIDSNNSLSEMTHERPOLHHSQDMVFTPESSGLQRLNEKLGTTAATELKKLDFKVSST	900
DB	841	PGAIDSNNSLSEMTHERPOLHHSQDMVFTPESSGLQRLNEKLGTTAATELKKLDFKVSST	900
QY	901	SNNLISTIPSDNLAAGTDNTSSLGPPSPMVHYQSDLTTLFGKSSPLTESGGPLSLEE	960
DB	901	SNNLISTIPSDNLAAGTDNTSSLGPPSPMVHYQSDLTTLFGKSSPLTESGGPLSLEE	960
QY	961	NNDKLLLESGLMNSQESSWGKNVSSSTESGRLFXGKRAHGPALLTKONALFKVSIILLKTN	1020
DB	961	NNDKLLLESGLMNSQESSWGKNVSSSTESGRLFXGKRAHGPALLTKONALFKVSIILLKTN	1020
QY	1021	KTSNNSATNRKTHIDGSPSLLIENSPSWQNLIESDTEFKVTPFLIHDRMLMDKNATALRL	1080
DB	1021	KTSNNSATNRKTHIDGSPSLLIENSPSWQNLIESDTEFKVTPFLIHDRMLMDKNATALRL	1080
QY	1081	NHNSKTTSSKNMVMQOKKEGPIPPDAQNPDMSFFKMLFLPESARWIOPTHKNSLNSG	1140
DB	1081	NHNSKTTSSKNMVMQOKKEGPIPPDAQNPDMSFFKMLFLPESARWIOPTHKNSLNSG	1140
QY	1141	QGESPKQLVSLGPEKVEGQNFLEKKNVYVVGGEFTKDVGLKEMVFPSSRNFLTNLDN	1200
DB	1141	QGESPKQLVSLGPEKVEGQNFLEKKNVYVVGGEFTKDVGLKEMVFPSSRNFLTNLDN	1200
QY	1201	LHENNTHNQKKTQEBLEKKTLLIOENVLPQTHVTGTGNFMKNLFLSTRQNVGSYD	1260
DB	1201	LHENNTHNQKKTQEBLEKKTLLIOENVLPQTHVTGTGNFMKNLFLSTRQNVGSYD	1260
QY	1261	GAYAPVLQDPRSLNDSTNRKTHAHFSKGBEENLEGLGNQTKQIVEKYACTTRISPN	1320
DB	1261	GAYAPVLQDPRSLNDSTNRKTHAHFSKGBEENLEGLGNQTKQIVEKYACTTRISPN	1320
QY	1321	SQNFVQSKRALQKFRIPLETELEKRIIVDDTSTQMSKNMKHITPSTLTQIDYNEKE	1380
DB	1321	SQNFVQSKRALQKFRIPLETELEKRIIVDDTSTQMSKNMKHITPSTLTQIDYNEKE	1380
QY	1381	KGAIOSPLSDCLTRSHSIPOANRSLPTAKVSSPFSIRPIYLRVLFQNSSHLPAASY	1440
DB	1381	KGAIOSPLSDCLTRSHSIPOANRSLPTAKVSSPFSIRPIYLRVLFQNSSHLPAASY	1440
QY	1441	RKXDSGVQSSSHFLQGAKKNNLSAILTLEMTDQREVSGLGTSATNSVYTKVENTVLP	1500
DB	1441	RKXDSGVQSSSHFLQGAKKNNLSAILTLEMTDQREVSGLGTSATNSVYTKVENTVLP	1500
QY	1501	KPOLPTXSGKVELLPKVHIYQKDLFPTETSNQSGPHLDLVEGSLQGTGEGAIKWNANRP	1560
DB	1501	KPOLPTXSGKVELLPKVHIYQKDLFPTETSNQSGPHLDLVEGSLQGTGEGAIKWNANRP	1560
QY	1561	GKVPFLAVATESAKTPSKLLDPLANDNHYGTQIPKEEWSQESPEKTAFFKDKTILSL	1620

Db 1561 GKVPFLRVATSSAKTPSKGLDPLAWDHYGTQLPKEWKSQEKSPKTAFFKXDDTILSL 1620
Qy 1621 NACESNHAIAINEQNKPELEVWAKOGRTERLCSQNPVLXKHQRHQRITRTTLQSDOEE 1680
Db 1621 NACESNHAIAINEQNKPELEVWAKOGRTERLCSQNPVLXKHQRHQRITRTTLQSDOEE 1680
Qy 1681 IDYDITSVENKKEFDYDDBENOSPRSFOKTRHYFIAAVERLWDYGMSSSPHVLNR 1740
Db 1681 IDYDITSVENKKEFDYDDBENOSPRSFOKTRHYFIAAVERLWDYGMSSSPHVLNR 1740
Qy 1741 AQSGSVPOFKVWFOEFTDGSFTQPLYGELNEHGLIGLPVIRAEVDENIMVTFNQASR 1800
Db 1741 AQSGSVPOFKVWFOEFTDGSFTQPLYGELNEHGLIGLPVIRAEVDENIMVTFNQASR 1800
Qy 1801 PYSFYSSLIISVEEDQOGAEPRKQFVKNEKTYFWKQVQHMAPTKDFDCKAWAYFSDV 1860
Db 1801 PYSFYSSLIISVEEDQOGAEPRKQFVKNEKTYFWKQVQHMAPTKDFDCKAWAYFSDV 1860
Qy 1861 DLEKDVHSLGILPILVCHTNILNPAHQVQVVOBPALEFTTIPDETCKSWYFTENMERNCR 1920
Db 1861 DLEKDVHSLGILPILVCHTNILNPAHQVQVVOBPALEFTTIPDETCKSWYFTENMERNCR 1920
Qy 1921 PCNQMEDPTKENTRFRAINGYIMDTLPLGLVMAQDQIRWVLLSMGSENIHSHFSGH 1980
Db 1921 PCNQMEDPTKENTRFRAINGYIMDTLPLGLVMAQDQIRWVLLSMGSENIHSHFSGH 1980
Qy 1981 VFTVRKKEEYKMALYNLPYGVFETVEMLPFSKAGIWRVECLIGELHAGMSTLFLVYSNKC 2040
Db 1981 VFTVRKKEEYKMALYNLPYGVFETVEMLPFSKAGIWRVECLIGELHAGMSTLFLVYSNKC 2040
Qy 2041 QTPLGMAQGHTRDFOITASGQYGWAPKARLHYSGSINANSTKPFPSWIKVDLLAPMII 2100
Db 2041 QTPLGMAQGHTRDFOITASGQYGWAPKARLHYSGSINANSTKPFPSWIKVDLLAPMII 2100
Qy 2101 HGKTCQARQKPFSSLIYISQFIIMYSLDGKKWOTYRGNSTGTLMVFGVNDSSGIXHNIFN 2160
Db 2101 HGKTCQARQKPFSSLIYISQFIIMYSLDGKKWOTYRGNSTGTLMVFGVNDSSGIXHNIFN 2160
Qy 2161 PPIIARYIRLHPTHYSIRSLAMELGMCDLNSCMLPGMESKAISDAQITASSYFTNMFA 2220
Db 2161 PPIIARYIRLHPTHYSIRSLAMELGMCDLNSCMLPGMESKAISDAQITASSYFTNMFA 2220
Qy 2221 TWSPSKARLHLOGRANRPQVNNKEMLOVDFOKTKVGTGTVGKSKLLTSMYVKEFL 2280
Db 2221 TWSPSKARLHLOGRANRPQVNNKEMLOVDFOKTKVGTGTVGKSKLLTSMYVKEFL 2280
Qy 2281 ISSSQDGHQWTLFFQNGKVKVFCGNQDSFTPVVNSLDPPILTRYLRIRHPQSWHQAIALRM 2340
Db 2281 ISSSQDGHQWTLFFQNGKVKVFCGNQDSFTPVVNSLDPPILTRYLRIRHPQSWHQAIALRM 2340
Qy 2341 EVLGCEAQDLY 2351
Db 2341 EVLGCEAQDLY 2351

RESULT 6
US-10-360-101-229
; Sequence 229, Application US/10360.01
; Publication No. US20040009550A1
; GENERAL INFORMATION:
; APPLICANT: Moll, Gert N.
; APPLICANT: Leenhouts, Cornelis J.
; TITLE OF INVENTION: Export and modification of (poly)peptide in the lantibiotic way
; FILE REFERENCE: 2183-5673
; CURRENT APPLICATION NUMBER: US/10/360.101
; CURRENT FILING DATE: 2003-02-07
; PRIOR APPLICATION NUMBER: EP 02077060.8
; PRIOR FILING DATE: 2002-05-24
; NUMBER OF SEQ ID NOS: 309
; SOFTWARE: Patent-In version 3.1
; SEQ ID NO 229
; LENGTH: 2332
; TYPE: PRT

; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: sequence of factor VIII
US-10-360-101-229

Query Match 99.0%; Score 12295; DB 15; Length 2332;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 2329; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 20 ATRRYIYGAVLSWDYMQSDLGELPVDARPPRPVKSPFPENTSVYKKTLPVBTDFHLEN 79
Db 1 ATRRYIYGAVLSWDYMQSDLGELPVDARPPRPVKSPFPENTSVYKKTLPVBTDFHLEN 60
Qy 80 IAKRPPRMWGLGPTIQAENVYDTVITLKNMASHPVSLHVGVSYWKASGAEDDQTSQ 139
Db 61 IAKRPPRMWGLGPTIQAENVYDTVITLKNMASHPVSLHVGVSYWKASGAEDDQTSQ 120
Qy 140 REKEDDKVFPGGSHTYVWQVLEKNGPMASDPLCLTYSLSHVDLVKDLNSGLIGALLVCR 199
Db 121 REKEDDKVFPGGSHTYVWQVLEKNGPMASDPLCLTYSLSHVDLVKDLNSGLIGALLVCR 180
Qy 200 EGSIAKEKTOHLKHPILLFAVFDGKSWHSETKSLMQDRDAASARAWPKMHTVNGYVNR 259
Db 181 EGSIAKEKTOHLKHPILLFAVFDGKSWHSETKSLMQDRDAASARAWPKMHTVNGYVNR 240
Qy 260 SLPLGIGCHRKSVVWHVIGMGTTPVHSIFLEHGTFLVRNHRQASLEISPTITFLTAQTL 319
Db 241 SLPLGIGCHRKSVVWHVIGMGTTPVHSIFLEHGTFLVRNHRQASLEISPTITFLTAQTL 300
Qy 320 MDLQOFLLSCHISHQHDGMEAYKVDSCPEEPQLIKKNNEAEYDDDLTDSMDVVRP 379
Db 301 MDLQOFLLSCHISHQHDGMEAYKVDSCPEEPQLIKKNNEAEYDDDLTDSMDVVRP 360
Qy 380 DDDKSPSPQIRSVAKKPKTWVHYIAAEEEDWDYAPLVLPAPDDRSYKSYQLNNGPQIRG 439
Db 361 DDDKSPSPQIRSVAKKPKTWVHYIAAEEEDWDYAPLVLPAPDDRSYKSYQLNNGPQIRG 420
Qy 440 RYKVKVRPMAYDTDTFTKTREAIQHESGILGPLLYGEVGDITLLIFKQASRPYNIYPHGI 499
Db 421 RYKVKVRPMAYDTDTFTKTREAIQHESGILGPLLYGEVGDITLLIFKQASRPYNIYPHGI 480
Qy 500 TDVRLYSRRIPKGVKHLKDFPILPGEIFKYKWTVTVEDGPTKSDPRCLTRYSSPVNME 559
Db 481 TDVRLYSRRIPKGVKHLKDFPILPGEIFKYKWTVTVEDGPTKSDPRCLTRYSSPVNME 540
Qy 560 RDLASGLIGPLLI CYKESVDQGNQIMSDKNVILFSVFDENRSMYLTENIQRLPNPAG 619
Db 541 RDLASGLIGPLLI CYKESVDQGNQIMSDKNVILFSVFDENRSMYLTENIQRLPNPAG 600
Qy 620 VQLEDPEFQASINIMHSINGYVFDLSQLSVCLHEVAYWILSIGAQTDPLSVFFSGYTFKH 679
Db 601 VQLEDPEFQASINIMHSINGYVFDLSQLSVCLHEVAYWILSIGAQTDPLSVFFSGYTFKH 660
Qy 680 KMWYEDTLTLPPFSGETVFMSENPGWLILGCHNSDPRNMGWTLALKVSSCDKNTGDIYE 739
Db 661 KMWYEDTLTLPPFSGETVFMSENPGWLILGCHNSDPRNMGWTLALKVSSCDKNTGDIYE 720
Qy 740 DSYEDISAYLLSKNAIEPRSFQSNRHPSTFQKFNATTPENDIEKTDZFAHRTMP 799
Db 721 DSYEDISAYLLSKNAIEPRSFQSNRHPSTFQKFNATTPENDIEKTDZFAHRTMP 780
Qy 800 KIQNVSSDILLMLLRQSPTHGLSLSDLOEAKYETFDSDPSGAIDSNNSLSETHFRPQ 859
Db 781 KIQNVSSDILLMLLRQSPTHGLSLSDLOEAKYETFDSDPSGAIDSNNSLSETHFRPQ 840
Qy 860 LHHSGDMVFTPESGQLRLNEKLGTTAATLKLKLDKFKVSSSTNNLISPTIPDNLAAGTDN 919
Db 841 LHHSGDMVFTPESGQLRLNEKLGTTAATLKLKLDKFKVSSSTNNLISPTIPDNLAAGTDN 900
Qy 920 TSSLGPPSPMPVHYLSQDLTTLFGKKSPLTBSGGPLSLSEENNDKSLLESGLMNSQESSW 979
Db 901 TSSLGPPSPMPVHYLSQDLTTLFGKKSPLTBSGGPLSLSEENNDKSLLESGLMNSQESSW 960

QY 960 GKNVSTESGRLFKGKRAHGPAALLTKDNLFAKVSISLLKTNKTSNNNSATNRKTHIDGPSL 1039
Db 961 GKNVSTESGRLFKGKRAHGPAALLTKDNLFAKVSISLLKTNKTSNNNSATNRKTHIDGPSL 1020
QY 1040 LIENSPTSVMQILLESDBTFFKQVTPLIHDBMLMDKNTALRINEMSNKTTSSKNEMVQOK 1099
Db 1021 LIENSPTSVMQILLESDBTFFKQVTPLIHDBMLMDKNTALRINEMSNKTTSSKNEMVQOK 1080
QY 1100 KEGPIPPDAQNPDMGFFKMLFLPESARWQRTGKNSNSGOGPSPKOLVSLGPEKSVEG 1159
Db 1081 KEGPIPPDAQNPDMGFFKMLFLPESARWQRTGKNSNSGOGPSPKOLVSLGPEKSVEG 1140
QY 1160 QNFLSEKXVVGKGEFTKDVGLKXNVFPSSRNLFNTNLDNLHNNHNNQKKEQEZIEK 1219
Db 1141 QNFLSEKXVVGKGEFTKDVGLKXNVFPSSRNLFNTNLDNLHNNHNNQKKEQEZIEK 1200
QY 1220 KETLIOENVLPOIHTVTGKNFMKNLFLLSPRQNVESYDCAVPLQDPRSLNDSNR 1279
Db 1201 KETLIOENVLPOIHTVTGKNFMKNLFLLSPRQNVESYDCAVPLQDPRSLNDSNR 1260
QY 1280 TKKHTAHFSKKEEBENLEGLNQTKQIVKQVACTTRISPNTSQQNFVQTKRALKQFRL 1339
Db 1261 TKKHTAHFSKKEEBENLEGLNQTKQIVKQVACTTRISPNTSQQNFVQTKRALKQFRL 1320
QY 1340 PLEETELEKRIIIVDDTSTQWKNMKHLPSTLTQIDYNEKXKGAITQSPLDCLTRSHSI 1399
Db 1321 PLEETELEKRIIIVDDTSTQWKNMKHLPSTLTQIDYNEKXKGAITQSPLDCLTRSHSI 1380
QY 1400 POANRSPPIAKVSFPSPRIPIYTRVLFOQNSHLPAAKSKDQSGVOESSHFLOGAKK 1459
Db 1381 POANRSPPIAKVSFPSPRIPIYTRVLFOQNSHLPAAKSKDQSGVOESSHFLOGAKK 1440
QY 1460 NNLSLAILTLEMTGQDREVGSLGTSATNSVTVKQVNTVLPKPDLPKTSKGVLELLPKVHI 1519
Db 1441 NNLSLAILTLEMTGQDREVGSLGTSATNSVTVKQVNTVLPKPDLPKTSKGVLELLPKVHI 1500
QY 1520 YOKOLPFTETNGSGPHLDLVEGSLLOQTGEGAIKNEANRPGKVPFLVATESSAKTSPK 1579
Db 1501 YOKOLPFTETNGSGPHLDLVEGSLLOQTGEGAIKNEANRPGKVPFLVATESSAKTSPK 1560
QY 1580 LLDPLAWNHYGTQIPKEEWSQKSPKTKKDTILSNACESHAIKAIINEGQKPK 1639
Db 1561 LLDPLAWNHYGTQIPKEEWSQKSPKTKKDTILSNACESHAIKAIINEGQKPK 1620
QY 1640 BIEVTWAKQGRTERICSONPPVLKREHREITRTTLQSDQEBIDYDDTISVMKXEDFDIY 1699
Db 1621 BIEVTWAKQGRTERICSONPPVLKREHREITRTTLQSDQEBIDYDDTISVMKXEDFDIY 1680
QY 1700 DEDENGSPRSFQKTRHYFIAAVERLWDYGMSSSPHVLNBAQSGSVDPQKXVFOQFTD 1759
Db 1681 DEDENGSPRSFQKTRHYFIAAVERLWDYGMSSSPHVLNBAQSGSVDPQKXVFOQFTD 1740
QY 1760 GSFTQPLRYGELNEHGLLGPYIRAEVEDNIMVTRNQAASPYFYSLISYEEDOROGA 1819
Db 1741 GSFTQPLRYGELNEHGLLGPYIRAEVEDNIMVTRNQAASPYFYSLISYEEDOROGA 1800
QY 1820 EPRNFVKNPNETKYFWKQVQHMAFTKDEPCKAWAYFSDVLEKDVHSGLIGLPLVCHT 1879
Db 1801 EPRNFVKNPNETKYFWKQVQHMAFTKDEPCKAWAYFSDVLEKDVHSGLIGLPLVCHT 1860
QY 1880 NTLNPAHGRQVTVQEFALFFTFIDETKSWYFETNMERNCRAPCNIQMEDPTFKENYRPHA 1939
Db 1861 NTLNPAHGRQVTVQEFALFFTFIDETKSWYFETNMERNCRAPCNIQMEDPTFKENYRPHA 1920
QY 1940 INGVIMDTPLGVVQAQDRIRWYLLSNGSNNIHSIHPSGHVFTVRKKEEYKALYNLYP 1999
Db 1921 INGVIMDTPLGVVQAQDRIRWYLLSNGSNNIHSIHPSGHVFTVRKKEEYKALYNLYP 1980
QY 2000 GVFTVEMLPKAGIWRVECLIGELHAGMSTLFLVYSNKKQCTPLGMASGHIRDFOITAS 2059
Db 1981 GVFTVEMLPKAGIWRVECLIGELHAGMSTLFLVYSNKKQCTPLGMASGHIRDFOITAS 2040
QY 2060 GOYQOWAPKLARLHYSGSINAWSKEPSPWIKVDLLAPMIHGIKTQAROKFSSJISQ 2119

Db 2041 GOYQOWAPKLARLHYSGSINAWSKEPSPWIKVDLLAPMIHGIKTQAROKFSSJISQ 2100
QY 2120 FIIMYSLDGKQWQYRGNSTGTLMVFFGNVDSSGFKNIFNPPIIARIYIRLHPHYSIRS 2179
Db 2101 FIIMYSLDGKQWQYRGNSTGTLMVFFGNVDSSGFKNIFNPPIIARIYIRLHPHYSIRS 2160
QY 2180 TLRVELMSCDNLNSCSMPLGMSKAISSAQITASSVFTNNFATWSPSKARLHLOGRSNAWR 2239
Db 2161 TLRVELMSCDNLNSCSMPLGMSKAISSAQITASSVFTNNFATWSPSKARLHLOGRSNAWR 2220
QY 2240 PQVANNPKEWLQVDQKTMKVTVGTTQGVKSLTSMYVKEFLISSQDGHQWTLFFONGKV 2299
Db 2221 PQVANNPKEWLQVDQKTMKVTVGTTQGVKSLTSMYVKEFLISSQDGHQWTLFFONGKV 2280
QY 2300 KYFOGQNDSPFPVNSLDPPLLTLYRLHRIHPOSVHQIALRMEVLGCEAQDLY 2351
Db 2281 KYFOGQNDSPFPVNSLDPPLLTLYRLHRIHPOSVHQIALRMEVLGCEAQDLY 2332
RESULT 7
US-09-957-641-2
; Sequence 2, Application US/09957641
; Publication No. US20020182670A1
; GENERAL INFORMATION:
; APPLICANT: Emory University
; TITLE OF INVENTION: MODIFIED FACTOR VIII
; FILE REFERENCE: 75-00
; CURRENT APPLICATION NUMBER: US/09/957,641
; CURRENT FILING DATE: 2001-09-16
; PRIOR APPLICATION NUMBER: US 60/234047
; PRIOR FILING DATE: 2000-09-19
; PRIOR APPLICATION NUMBER: US 60/236460
; PRIOR FILING DATE: 2000-09-25
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 2332
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-957-641-2
Query Match 98.9%; Score 12282; DB 9; Length 2332;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 2327; Conservative 2; Mismatches 3; Indels 0; Gaps 0;
QY 20 ATRRYYLGAVELSDWYMQSDLGELPVDARPPRPVKSPFPNTSVVYKTLFVEPTHLEN 79
Db 1 ATRRYYLGAVELSDWYMQSDLGELPVDARPPRPVKSPFPNTSVVYKTLFVEPTHLEN 60
QY 80 IAKPRPFWMLGPTIOAEVVDVTWITLKNWASHPVSLHAGVSVYKASGAEVDDOTSQ 139
Db 61 IAKPRPFWMLGPTIOAEVVDVTWITLKNWASHPVSLHAGVSVYKASGAEVDDOTSQ 120
QY 140 REKEDDKVFPQGSHTYVQVLEKNGPMASDPLCTYSYLSHVDLVKDLNSGLICALLVCR 199
Db 121 REKEDDKVFPQGSHTYVQVLEKNGPMASDPLCTYSYLSHVDLVKDLNSGLICALLVCR 180
QY 200 EGSIAKEKTQTLHKFILLFAVFBEGKSWHSETKNSLMODDAASARAWPKMHTVNGVNR 259
Db 181 EGSIAKEKTQTLHKFILLFAVFBEGKSWHSETKNSLMODDAASARAWPKMHTVNGVNR 240
QY 260 SLPLGLICHRKSVVHVIGMTTPEVHSIFLEGHTFLVRNHRQASLSIPTITLTAOTLL 319
Db 241 SLPLGLICHRKSVVHVIGMTTPEVHSIFLEGHTFLVRNHRQASLSIPTITLTAOTLL 300
QY 320 MDLQGFLLSCHISQHDGMEAYVKVDSCEPEPQLIMKNEEAEDYDDDLTDSMDVVR 379
Db 301 MDLQGFLLSCHISQHDGMEAYVKVDSCEPEPQLIMKNEEAEDYDDDLTDSMDVVR 360
QY 380 DDNSPSFPIQIRSVAKKHPKTHWHYIAAEEEDWDYAPLAPDDRSYKSYLNNGPORIG 439
Db 361 DDNSPSFPIQIRSVAKKHPKTHWHYIAAEEEDWDYAPLAPDDRSYKSYLNNGPORIG 420

QY 440 RYKXVRFMAYTDETFKTRAEAIQHESGILGPLYGEVGDITLIIIFKNQASRPYNTPHGI 499
 DB 42- RYKXVRFMAYTDETFKTRAEAIQHESGILGPLYGEVGDITLIIIFKNQASRPYNTPHGI 480
 QY 500 TVDRPLYSRLSGVGHLLKDFILPQEPKYKWTVTVEGPKTSDPRCLTRYYSFVNM 559
 DB 481 TVDRPLYSRLSGVGHLLKDFILPQEPKYKWTVTVEGPKTSDPRCLTRYYSFVNM 540
 QY 560 RLASGLIPLIICYKESVDQXGNOIMSKRNVILFSVFDENRSVLTENIORFENPAG 619
 DB 541 RLASGLIPLIICYKESVDQXGNOIMSKRNVILFSVFDENRSVLTENIORFENPAG 600
 QY 620 VOLEPPEQASIMESINGYVFDLSQLSVCLHEVAYWYILSIGAOTDFLSVFFSGVTPKH 679
 DB 601 VOLEPPEQASIMESINGYVFDLSQLSVCLHEVAYWYILSIGAOTDFLSVFFSGVTPKH 660
 QY 680 KMYVEDTLTLFPFSGETVPMSENEPCLWILGCHNSDFRNGMTALLKVSSCDKNTGDYIE 739
 DB 661 KMYVEDTLTLFPFSGETVPMSENEPCLWILGCHNSDFRNGMTALLKVSSCDKNTGDYIE 720
 QY 740 DSYEDISAYLLSKNNAIPRSPSONSRHPSRTPKOFNATITIPENDIEKTDPMFAHRTMP 799
 DB 721 DSYEDISAYLLSKNNAIPRSPSONSRHPSRTPKOFNATITIPENDIEKTDPMFAHRTMP 780
 QY 800 KTONVSSDMLLRQSPTPHGLSLSDIQEAKYETPSDDPSGCAIDSNNSLSEMTFRPQ 859
 DB 781 KTONVSSDMLLRQSPTPHGLSLSDIQEAKYETPSDDPSGCAIDSNNSLSEMTFRPQ 840
 QY 860 LHSQDMVFTPESGLOLRNEKLTGTAATELAKLPKYSSTSNNLSTIPSDNLAAGTDN 919
 DB 841 LHSQDMVFTPESGLOLRNEKLTGTAATELAKLPKYSSTSNNLSTIPSDNLAAGTDN 900
 QY 920 TSSLGPPMPVHYDQDITLTPGKSSPLTESGGPLSLEENNDKLLSGLMNSQESSW 979
 DB 901 TSSLGPPMPVHYDQDITLTPGKSSPLTESGGPLSLEENNDKLLSGLMNSQESSW 960
 QY 980 GKNVSTSGRLFKGRAGPALLTKDNALFKVSISLLKTNKTSNNSATNRKTHIDGPSL 1039
 DB 961 GKNVSTSGRLFKGRAGPALLTKDNALFKVSISLLKTNKTSNNSATNRKTHIDGPSL 1020
 QY 1040 LIENSPSVQNLISDTEFKVTPLIHORMMDKNATA:RLNHSNKYKTSKKNEMVQOK 1099
 DB 1021 LIENSPSVQNLISDTEFKVTPLIHORMMDKNATA:RLNHSNKYKTSKKNEMVQOK 1080
 QY 1100 KEGPIPPDAQNPDXSFFKMLFLPESARMIQRTGKNSLNSGGPSPKQVLSLGPESKSV 1159
 DB 1081 KEGPIPPDAQNPDXSFFKMLFLPESARMIQRTGKNSLNSGGPSPKQVLSLGPESKSV 1140
 QY 1160 QNPLSEKNKVVGKEFTKDVGLKEMVFPSSRNLFITNLDNLHNNNTNQEKKIOEETEK 1219
 DB 1141 QNPLSEKNKVVGKEFTKDVGLKEMVFPSSRNLFITNLDNLHNNNTNQEKKIOEETEK 1200
 QY 1220 KETLTOENVLPQIHTVTGKTGNFMKNLFLLSRQNVESYDGAAPV:QDPSLNDSTNR 1279
 DB 1201 KETLTOENVLPQIHTVTGKTGNFMKNLFLLSRQNVESYDGAAPV:QDPSLNDSTNR 1260
 QY 1280 TKKHTAHSKKEENLEBGLNQTKQIWEKYACTTRISNTSQQNFVTRSKRALKQRL 1339
 DB 1261 TKKHTAHSKKEENLEBGLNQTKQIWEKYACTTRISNTSQQNFVTRSKRALKQRL 1320
 QY 1340 PLEETELEKRIIVDDTQWSKNMGLTBSLTQIDYNEKEKGALITQSPDCLTRSHSI 1399
 DB 1321 PLEETELEKRIIVDDTQWSKNMGLTBSLTQIDYNEKEKGALITQSPDCLTRSHSI 1380
 QY 1400 POANRSPPLIAKVSFPPIRPIYLRVLFQDNSHLPAAVSFKDQSGVQESSHFLQGAKK 1459
 DB 1381 POANRSPPLIAKVSFPPIRPIYLRVLFQDNSHLPAAVSFKDQSGVQESSHFLQGAKK 1440
 QY 1460 NNLSLAITLJEMTGDOREVGSLGTSATNSVTYKVKVENTVLPKPDLPKTSKGVELLPKVHI 1519
 DB 1441 NNLSLAITLJEMTGDOREVGSLGTSATNSVTYKVKVENTVLPKPDLPKTSKGVELLPKVHI 1500

QY 1520 YOKOLFPPTETSGSPGHLDLVEGSLLOQTEGAIKWNEANRPGKVPFLRVATESSAKTSPK 1579
 DB 1501 YOKOLFPPTETSGSPGHLDLVEGSLLOQTEGAIKWNEANRPGKVPFLRVATESSAKTSPK 1560
 QY 1580 LLDPLANDNHVGTQIPEKEWSQESKPEKTAFAKKDITILSNACESNHAIAINEGQNK 1639
 DB 1561 LLDPLANDNHVGTQIPEKEWSQESKPEKTAFAKKDITILSNACESNHAIAINEGQNK 1620
 QY 1640 EIEVTWAKQGTERTLCSQNPVLAKEHOREITRTTLOSQOEIDYDDTISVEMKEDFDIY 1699
 DB 1621 EIEVTWAKQGTERTLCSQNPVLAKEHOREITRTTLOSQOEIDYDDTISVEMKEDFDIY 1680
 QY 1700 DEDENQSPRSQKKTTRHYFIAAVERLWDYGMSSPHVLNRNAQSGSVPOFKVWVFOETD 1759
 DB 1681 DEDENQSPRSQKKTTRHYFIAAVERLWDYGMSSPHVLNRNAQSGSVPOFKVWVFOETD 1740
 QY 1760 GSFTQPIYRGELNHLGLLGPYIRAEVEDNIMVTRNOASRPYFYSLSIYEBEDOROGA 1819
 DB 1741 GSFTQPIYRGELNHLGLLGPYIRAEVEDNIMVTRNOASRPYFYSLSIYEBEDOROGA 1800
 QY 1820 EPRKNFVKPNETKTYFKVVOHHMPTKDEPCKKAWAYPSDVLKDVHSGLIGLLVCHT 1879
 DB 1801 EPRKNFVKPNETKTYFKVVOHHMPTKDEPCKKAWAYPSDVLKDVHSGLIGLLVCHT 1860
 QY 1880 NTLAPAGHQRTVQOEFALEFFITFDETKSWYFTENMERNCRAPCNIOEMDPTFKENYRPHA 1939
 DB 1861 NTLAPAGHQRTVQOEFALEFFITFDETKSWYFTENMERNCRAPCNIOEMDPTFKENYRPHA 1920
 QY 1940 INGIMDTPLGLVNAQDORIIRWYLLSGNSNENIHSIHPSGHVFTVRKKEEYKMALYNLYP 1999
 DB 1921 INGIMDTPLGLVNAQDORIIRWYLLSGNSNENIHSIHPSGHVFTVRKKEEYKMALYNLYP 1980
 QY 2000 GVETEMVLPKAGIWRVECLIGHLHAGMSTLFLVYSNKCOTPLGMASCHIRPOFITAS 2059
 DB 1981 GVETEMVLPKAGIWRVECLIGHLHAGMSTLFLVYSNKCOTPLGMASCHIRPOFITAS 2040
 QY 2060 GOYQWAPKLAARLHYSGSINAWSTKEPPSWIKVOLLAPMIHGIKTQAROKFSSLYISQ 2119
 DB 2041 GOYQWAPKLAARLHYSGSINAWSTKEPPSWIKVOLLAPMIHGIKTQAROKFSSLYISQ 2100
 QY 2120 FIIMYSLDGKKWQYRGNSTGTLVFFGNVDSGGIKNINFPDPIIARYIRLHPHYSIRS 2179
 DB 2101 FIIMYSLDGKKWQYRGNSTGTLVFFGNVDSGGIKNINFPDPIIARYIRLHPHYSIRS 2160
 QY 2180 TLRVLMGCDLNSCMLPGMBSKAIQAQITASSYFTNMPATWSPSKARLLHOGRSNAWR 2239
 DB 2161 TLRVLMGCDLNSCMLPGMBSKAIQAQITASSYFTNMPATWSPSKARLLHOGRSNAWR 2220
 QY 2240 PQVNNPKEMLOVDFOKTKKVTGTQGVKSLLTSMYVKEFLISSQDGHQWTLFPQNGKV 2299
 DB 2221 PQVNNPKEMLOVDFOKTKKVTGTQGVKSLLTSMYVKEFLISSQDGHQWTLFPQNGKV 2280
 QY 2300 KVFQGNQDSFTVNSLDPPLTRYLRIHPQSWHQTALRMEVLGCEAQDLY 2351
 DB 2281 KVFQGNQDSFTVNSLDPPLTRYLRIHPQSWHQTALRMEVLGCEAQDLY 2332

RESULT 8

US-10-187-319-2
 ; Sequence 2, Application US/10187319
 ; Publication No. US2003006878SAI
 ; GENERAL INFORMATION:
 ; APPLICANT: Lollar, John S.
 ; TITLE OF INVENTION: Hybrid Human/Animal Factor VIII
 ; NUMBER OF SEQUENCES: 40
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Greenlee, Winner and Sullivan, P.C.
 ; STREET: 5370 Manhattan Circle Suite 201
 ; CITY: Boulder
 ; STATE: Colorado
 ; COUNTRY: USA
 ; ZIP: 80303
 ; COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC Compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patentin Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/10/187,319
 FILING DATE: 27-Aug-2002
 CLASSIFICATION: <Unknown>
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 09/523,656
 FILING DATE: 2000-03-10
 APPLICATION NUMBER: US 09/037,601
 FILING DATE: 1998-03-10
 APPLICATION NUMBER: WO PCT/US97/11155
 FILING DATE: 1997-06-26
 APPLICATION NUMBER: US 06/670,707
 FILING DATE: 1996-06-26
 ATTORNEY/AGENT INFORMATION:
 NAME: Greenlee, Lorance L.
 REGISTRATION NUMBER: 27,894
 REFERENCE/DOCKET NUMBER: 75-95K
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 303/499-8080
 TELEFAX: 303/499-8089
 INFORMATION FOR SEQ ID NO: 2:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 2332 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: <Unknown>
 MOLECULE TYPE: protein
 HYPOTHETICAL: YES
 ANTI-SENSE: NO
 FRAGMENT TYPE: N-terminal
 ORIGINAL SOURCE:
 ORGANISM: Homo sapiens
 TISSUE TYPE: Liver
 SEQUENCE DESCRIPTION: SEQ ID NO: 2:
 US-10-187-319-2

Query Match 98.98; Score 12282; DB 14; Length 2332;
 Best Local Similarity 99.88; Pred. No. 0;
 Matches 232; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY	20	ATRRYILGAVELSDYMQSDLGELPVDARPPRPVKSFPENTSVVYKTLFVFTDHLFN	79
DB	1	ATRRYILGAVELSDYMQSDLGELPVDARPPRPVKSFPENTSVVYKTLFVFTVHLFN	60
QY	80	IAKPRPPWMLLGPFTQAEVYDVTWITLKNASHPVSLHAGVSYWKASGCAEYDDQTSQ	139
DB	61	IAKPRPPWMLLGPFTQAEVYDVTWITLKNASHPVSLHAGVSYWKASGCAEYDDQTSQ	120
QY	140	REKEDKVPFGSGHTYVQVLKENGPMASDPLCLITYSLSHVDLVLDLNSGLLGALLVCR	199
DB	121	REKEDKVPFGSGHTYVQVLKENGPMASDPLCLITYSLSHVDLVLDLNSGLLGALLVCR	180
QY	200	EGSLAKKTKTLHKFILLFAVDEGKSWHSETKNSLMQDRDAASARAWPQHTWNGVYNR	259
DB	181	EGSLAKKTKTLHKFILLFAVDEGKSWHSETKNSLMQDRDAASARAWPQHTWNGVYNR	240
QY	260	SLPGLGCHRSKYVWVIGMTTPEVHSIFLEGHTFLVRNHRQASLEISPTIFLTAQTLL	319
DB	241	SLPGLGCHRSKYVWVIGMTTPEVHSIFLEGHTFLVRNHRQASLEISPTIFLTAQTLL	300
QY	320	MDLGQFLLSCHISSHQHDGMEAYVKVDSCPEEPQLIMKNEEAEYDDDLTDSMDVVRP	379
DB	301	MDLGQFLLPCHISSHQHDGMEAYVKVDSCPEEPQLIMKNEEAEYDDDLTDSMDVVRP	360
QY	380	DDNSPFSFIQIRSVAKKHPKTHVHYIAAEDEDWDYAPLVAPDDRYSKSOYLNNGPQIRG	439
DB	361	DDNSPFSFIQIRSVAKKHPKTHVHYIAAEDEDWDYAPLVAPDDRYSKSOYLNNGPQIRG	420
QY	440	RKYKVRFMAYTDETPKTRAIQHESGILGPLLYGEVGDTLILLIPKMQASRPNIYPHGI	499

DB	421	RKYKVRFMAYTDETPKTRAIQHESGILGPLLYGEVGDTLILLIPKMQASRPNIYPHGI	480
QY	500	TDVRPLYSRRLPKGVKHLKDFPILPGEIFKYKVTVTVEDGFTKSDPRCLTRYSSVFVME	559
DB	481	TDVRPLYSRRLPKGVKHLKDFPILPGEIFKYKVTVTVEDGFTKSDPRCLTRYSSVFVME	540
QY	560	RDLASGLIGELLICYKESVDOKNGNOIMSDKRNVLFSVFDENRSWYLTENIQFLNPAG	619
DB	541	RDLASGLIGELLICYKESVDQNGNQIMSDKRNVLFSVFDENRSWYLTENIQFLNPAG	600
QY	620	VOLEDPEFQASNMHMSINGVYVDSLOL SVCLHEVAYWYILSIGAQTDFLVPFSGYTFKH	679
DB	601	VOLEDPEFQASNMHMSINGVYVDSLOL SVCLHEVAYWYILSIGAQTDFLVPFSGYTFKH	660
QY	680	KMYVEDTLILFPFSGETVYKSMENPGLWILGCHNSDPNRGMPTALLKVSSCDKXNTGYE	739
DB	661	KMYVEDTLILFPFSGETVYKSMENPGLWILGCHNSDPNRGMPTALLKVSSCDKXNTGYE	720
QY	740	DSYEDISAYLLSKNNAIEPRSPNSRHPSTRQKQFNATTIPENDIEKTPWFAHRTMP	799
DB	721	DSYEDISAYLLSKNNAIEPRSPNSRHPSTRQKQFNATTIPENDIEKTPWFAHRTMP	780
QY	800	KIQNVSSDILLMLLRQSPTPHGLSLSDLOAKYETESDDPSGAISSNNSLSEMTFRPQ	859
DB	781	KIQNVSSDILLMLLRQSPTPHGLSLSDLOAKYETESDDPSGAISSNNSLSEMTFRPQ	840
QY	860	LHSGDMVFTPEGLOLRNLKLTGTTAATLKKLDPKVSSTSNKLISTIPSDMLAAGTDN	919
DB	841	LHSGDMVFTPEGLOLRNLKLTGTTAATLKKLDPKVSSTSNKLISTIPSDMLAAGTDN	900
QY	920	TSSLGPPSMPVHYDSQDITLFGKKSPTESGSPISLSENNDSKLLSGLMNSQESSW	979
DB	901	TSSLGPPSMPVHYDSQDITLFGKKSPTESGSPISLSENNDSKLLSGLMNSQESSW	960
QY	980	GKNVSTESGRLPFKGRAHGPALLTKDNALFKVSI SLLTKNTSNNSATNRKTHIDGPSL	1039
DB	961	GKNVSTESGRLPFKGRAHGPALLTKDNALFKVSI SLLTKNTSNNSATNRKTHIDGPSL	1020
QY	1040	LIENSFVQWNLLESDETEPKKVTPLIHDRMLMDKNATLRLNMSNKTTSKXNMVQOK	1099
DB	1021	LIENSFVQWNLLESDETEPKKVTPLIHDRMLMDKNATLRLNMSNKTTSKXNMVQOK	1080
QY	1100	KEGPIPPDAQNPDMSPFKMLFLPESARWQRTGKNSLNSGGQSPKQVLSLGPESKVEG	1159
DB	1081	KEGPIPPDAQNPDMSPFKMLFLPESARWQRTGKNSLNSGGQSPKQVLSLGPESKVEG	1140
QY	1160	QNFLSEKKNVYVKGKFTKDVGLKEMVFPSSRNFLTLNLDNLHNNTHNQEKIQEIEK	1219
DB	1141	QNFLSEKKNVYVKGKFTKDVGLKEMVFPSSRNFLTLNLDNLHNNTHNQEKIQEIEK	1200
QY	1220	KETLIOENVVLQIHTVTGTGKPMKNLFLLSVQNVGSGVAGAYLQDFSLNDSTNR	1279
DB	1201	KETLIOENVVLQIHTVTGTGKPMKNLFLLSVQNVGSGVAGAYLQDFSLNDSTNR	1260
QY	1280	TKKHTAHFKKGBEENLEGLGNOTKQIVKAYCTTRISNTSQNFVTOQSRKALKQFRL	1339
DB	1261	TKKHTAHFKKGBEENLEGLGNOTKQIVKAYCTTRISNTSQNFVTOQSRKALKQFRL	1320
QY	1340	PLEETELEXRIIVDDTSTQWSKNMKHLTSTLTQIDYNEKEKGAITQSPSLDCLTRSHSI	1399
DB	1321	PLEETELEXRIIVDDTSTQWSKNMKHLTSTLTQIDYNEKEKGAITQSPSLDCLTRSHSI	1380
QY	1400	POANRSPLEIAKVSFPSPRIPLFVLFDQNSHLPAASYKDKSCGVQESSHFLQCAKK	1459
DB	1381	POANRSPLEIAKVSFPSPRIPLFVLFDQNSHLPAASYKDKSCGVQESSHFLQCAKK	1440
QY	1460	NNLSAILILEMTGOREVYGLGTATNNTVTKVENTVLPKPDLPKXTSGKVELLPKVHI	1519
DB	1441	NNLSAILILEMTGOREVYGLGTATNNTVTKVENTVLPKPDLPKXTSGKVELLPKVHI	1500
QY	1520	YOKDLPTFTSNGSFEGLDLVEGSLQGTGEGAIKWNEANRPGKVPFLRVATESSAKTPSK	1579

Db 1501 YQKLLPPTTSNGSGHLDLVEGSLLOETBGAIKWNEANRPGKVPFLRVATESSAKTPSK 1560
Qy 1580 LLDPLAKNHHYGTQPKBEWSQBSPEKTAPKKTOTILSNACSNHAIAINEGQKPK 1639
Db 1561 LLDPLAKNHHYGTQPKBEWSQBSPEKTAPKKTOTILSNACSNHAIAINEGQKPK 1620
Qy 1640 EIEVTAQGRTERCSQNPVPLKHHQREIIFTTTQSQQEIDYDDTISVEMKXEDFIY 1699
Db 1621 EIEVTAQGRTERCSQNPVPLKHHQREIIFTTTQSQQEIDYDDTISVEMKXEDFIY 1680
Qy 1700 DEDENQSRFQKTRHYPIAVERLWYGMSSSHVLRNRAQSGSVPOFKKVPQEBTD 1759
Db 1681 DEDENQSRFQKTRHYPIAVERLWYGMSSSHVLRNRAQSGSVPOFKKVPQEBTD 1740
Qy 1760 CSFTQPLVGLNEHLGLLGPVIRAEVDNIMVTRNQASRPSYSSLSIYEEBQROGA 1819
Db 1741 CSFTQPLVGLNEHLGLLGPVIRAEVDNIMVTRNQASRPSYSSLSIYEEBQROGA 1800
Qy 1820 EPRKNFVAPNETKTYFWKQHEMAPTKDEPCKAWAYFSDVDLEKDVASGLIGPLLVCHT 1879
Db 1801 EPRKNFVAPNETKTYFWKQHEMAPTKDEPCKAWAYFSDVDLEKDVASGLIGPLLVCHT 1860
Qy 1880 NTLNFAHGRQVTVQBFALFFITFDETKSWYFTENNERCRAPCNQMEDPTFKENYRFA 1939
Db 1861 NTLNFAHGRQVTVQBFALFFITFDETKSWYFTENNERCRAPCNQMEDPTFKENYRFA 1920
Qy 1940 INGYMDTLPGLVMAQDORIRWILLSMGSENIHSHFSGHVFTYRKKEEKWALYNLYP 1999
Db 1921 INGYMDTLPGLVMAQDORIRWILLSMGSENIHSHFSGHVFTYRKKEEKWALYNLYP 1980
Qy 2000 GYFTEVEMLPKAGIWRVECLIGEHLHAGMSTFLVYGNKCTPLGMASGHIRDPQITAS 2059
Db 1981 GYFTEVEMLPKAGIWRVECLIGEHLHAGMSTFLVYGNKCTPLGMASGHIRDPQITAS 2040
Qy 2060 QYQGWAPKARLHVSGSINAWSTKEPSWIKVDLLAPMIHGIKTQARQKFSLSYISQ 2119
Db 2041 QYQGWAPKARLHVSGSINAWSTKEPSWIKVDLLAPMIHGIKTQARQKFSLSYISQ 2100
Qy 2120 FIIMYSLGCKKWQTRGNSTGLTVFFGNVDSSGKKNHFNPIIARVIRLHPHYSIRS 2179
Db 2101 FIIMYSLGCKKWQTRGNSTGLTVFFGNVDSSGKKNHFNPIIARVIRLHPHYSIRS 2160
Qy 2180 TIRBELMGCDLNSCNSMPLGMESKASDAQITASSYFTNMFTWSPSKARLHLQGRSNWR 2239
Db 2161 TIRBELMGCDLNSCNSMPLGMESKASDAQITASSYFTNMFTWSPSKARLHLQGRSNWR 2220
Qy 2240 PQVNNPKELQVDFOKTKVQVTTQGYKSLTSMYVKEFLISSQDGHQWTLFPQNGKV 2299
Db 2221 PQVNNPKELQVDFOKTKVQVTTQGYKSLTSMYVKEFLISSQDGHQWTLFPQNGKV 2280
Qy 2300 KVFOGNQDSFTPVVNSLDPPLTRYLRIRHPQSVWHQIALRKEVLCGCAQDLY 2351
Db 2281 KVFOGNQDSFTPVVNSLDPPLTRYLRIRHPQSVWHQIALRKEVLCGCAQDLY 2332

RESULT 9

US-10-131-510A-2
; Sequence 2, Application US/10131510A
; Publication No. US20030166536A1
; GENERAL INFORMATION:
; APPLICANT: Jollard, John S
; TITLE OF INVENTION: Modified Factor VII
; FILE REFERENCE: 75-95J
; CURRENT APPLICATION NUMBER: US/10/131,510A
; CURRENT FILING DATE: 2002-07-10
; PRIOR APPLICATION NUMBER: U.S. 09/315,179
; PRIOR FILING DATE: 1999-05-20
; PRIOR APPLICATION NUMBER: U.S. 09/037,601
; PRIOR FILING DATE: 1998-03-10
; PRIOR APPLICATION NUMBER: U.S. 08/670,707
; PRIOR FILING DATE: 1996-06-26
; PRIOR APPLICATION NUMBER: PCT/US97/11155
; PRIOR FILING DATE: 1997-06-26

; PRIOR APPLICATION NUMBER: PCT/US94/13200
; PRIOR FILING DATE: 1994-11-15
; PRIOR APPLICATION NUMBER: U.S. 08/212,133
; PRIOR FILING DATE: 1994-03-11
; PRIOR APPLICATION NUMBER: U.S. 07/864,004
; PRIOR FILING DATE: 1992-04-07
; NUMBER OF SEQ ID NOS: 40
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 2332
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-131-510A-2

Query Match 98.9%; Score 12282; DB 14; Length 2332;

Best Local Similarity 99.8%; Pred. No. 0;

Matches 2327; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Qy 20 ATRRYYLCAVELSDYMQSDLGELPVDARPPRPVKSPFNTSVYKXTLFFVEFTHLEN 79
Db 1 ATRRYYLCAVELSDYMQSDLGELPVDARPPRPVKSPFNTSVYKXTLFFVEFTHLEN 60
Qy 80 IAXRPPPMWGLLGTIIQAEVYDTVITLKNMASHPVSLHVGVSYWKASGEAEYDDQTSQ 139
Db 61 IAXRPPPMWGLLGTIIQAEVYDTVITLKNMASHPVSLHVGVSYWKASGEAEYDDQTSQ 120
Qy 140 REKEDDKVPPGSGSHYVQVILKENGXASDPCLITYSYLSHVDLVKDLNSGLIGALLVCR 199
Db 121 REKEDDKVPPGSGSHYVQVILKENGXASDPCLITYSYLSHVDLVKDLNSGLIGALLVCR 180
Qy 200 EGSLLAKEKXTQLHFKFILLFAVFDGKSWHSTKNSLMQDRDAASARAWPKMHTVNGVNR 259
Db 181 EGSLLAKEKXTQLHFKFILLFAVFDGKSWHSTKNSLMQDRDAASARAWPKMHTVNGVNR 240
Qy 260 SLPLGLIGHRKSVYVHVGIMGTTPZVHSIFLEGHTFLVRNHRQASLSIPTFLTAQTL 319
Db 241 SLPLGLIGHRKSVYVHVGIMGTTPZVHSIFLEGHTFLVRNHRQASLSIPTFLTAQTL 300
Qy 320 MDLQGLLSCHISGHQDGMAYKVDSCPEPOLIKKNNEAEADYDDDLTDSMDVVRP 379
Db 301 MDLQGLLSCHISGHQDGMAYKVDSCPEPOLIKKNNEAEADYDDDLTDSMDVVRP 360
Qy 380 DDDNSPFIQIRSVAKKHPKTVWHYIAAEEDWDYAPLVLPADDERSYKSOYLNNGPQIG 439
Db 361 DDDNSPFIQIRSVAKKHPKTVWHYIAAEEDWDYAPLVLPADDERSYKSOYLNNGPQIG 420
Qy 440 RYKVKRPMAYTDETFKTRAIQHESGILGPLLYGEVGDITLLIFKQASRPYNIYPHGI 499
Db 421 RYKVKRPMAYTDETFKTRAIQHESGILGPLLYGEVGDITLLIFKQASRPYNIYPHGI 480
Qy 500 TDVRPLYSRRLPKGVKHLKDPPLPGEIFKYKWTVTVEDGPTKSDPRCLTLYSSFVNME 559
Db 481 TDVRPLYSRRLPKGVKHLKDPPLPGEIFKYKWTVTVEDGPTKSDPRCLTLYSSFVNME 540
Qy 560 RDLASGLTGPLLCYKSDQGNQIMSDKENVILFSVDENRSMYLTENIQRELPNPAG 619
Db 541 RDLASGLTGPLLCYKSDQGNQIMSDKENVILFSVDENRSMYLTENIQRELPNPAG 600
Qy 620 VQLEDPEFQASNIMHSGYVDSLSQSVCLHEVAYWYILSIGAQTFLSVFFSGYTFKH 679
Db 601 VQLEDPEFQASNIMHSGYVDSLSQSVCLHEVAYWYILSIGAQTFLSVFFSGYTFKH 660
Qy 680 KMVYEDTLTLPFPGSETVFMENPGLWILGCHNSDFRNRGMTALLKVSXCDKXGTGYE 739
Db 661 KMVYEDTLTLPFPGSETVFMENPGLWILGCHNSDFRNRGMTALLKVSXCDKXGTGYE 720
Qy 740 DSYEDISAYLSKNAIEPRFSQNSRHPSTRQKQFNATTIPENDIEKTDWFAHRTMP 799
Db 721 DSYEDISAYLSKNAIEPRFSQNSRHPSTRQKQFNATTIPENDIEKTDWFAHRTMP 780
Qy 800 KIQNVSSDDLMLLRQSPTHGLSLSLQAEAKYTFPSDDPSGAIDNNSLSETHFRPQ 859
Db 781 KIQNVSSDDLMLLRQSPTHGLSLSLQAEAKYTFPSDDPSGAIDNNSLSETHFRPQ 840

Qy 860 LHHSGDMVFTPEESGLQRLNEKLGTTAATELKKLDPKVSSSTNNLISITIPSDNLAAGTDN 919
Db 841 LHHSGDMVFTPEESGLQRLNEKLGTTAATELKKLDPKVSSSTNNLISITIPSDNLAAGTDN 900
Qy 920 TSSIGPPMPVHYDSQDITTTGKSSPLTBSGGPLSSEENNDKLESGLMNSQESSW 979
Db 901 TSSIGPPMPVHYDSQDITTTGKSSPLTBSGGPLSSEENNDKLESGLMNSQESSW 960
Qy 980 GKNYSSSTESGRLFKGRAGHALLTKDNALPKVSIISLTKTKTSNNSTNKTHTIDGSL 1039
Db 961 GKNYSSSTESGRLFKGRAGHALLTKDNALPKVSIISLTKTKTSNNSTNKTHTIDGSL 1020
Qy 1040 LIENSPPVQNLISBDETFKVTPLIHDRMLMDKNATRLNHNMSNKTSSKNMEMVOOK 1099
Db 1021 LIENSPPVQNLISBDETFKVTPLIHDRMLMDKNATRLNHNMSNKTSSKNMEMVOOK 1080
Qy 1100 KEGTIPPPDAQNPDMSFFKMLFLPESARMTQTHGKNSLNSGGPPKQLVSLGPBKSVEG 1159
Db 1081 KEGTIPPPDAQNPDMSFFKMLFLPESARMTQTHGKNSLNSGGPPKQLVSLGPBKSVEG 1140
Qy 1160 QNFISEKKNVVGKEFTKDYGLKEMVFPSSNLEFLTNLDNLHNNTHNQEKKEIEK 1219
Db 1141 QNFISEKKNVVGKEFTKDYGLKEMVFPSSNLEFLTNLDNLHNNTHNQEKKEIEK 1200
Qy 1220 KETLIQENNVLPQIHTVTGTRKNFMKNLFLLSRQNVESYDGAAPVLQDPRESLNDSTNR 1279
Db 1201 KETLIQENNVLPQIHTVTGTRKNFMKNLFLLSRQNVESYDGAAPVLQDPRESLNDSTNR 1260
Qy 1280 TKKHTAHKSKGEBEENLSEGLNCTQIIVKXACTTRISPNTSQQNFVQTSKRALKQRL 1339
Db 1261 TKKHTAHKSKGEBEENLSEGLNCTQIIVKXACTTRISPNTSQQNFVQTSKRALKQRL 1320
Qy 1340 PLEETELEKRIIIVDTSTQWSKNMKHLTPSTLTQIDYNEKEKGAITQSPSLCLTRSHI 1399
Db 1321 PLEETELEKRIIIVDTSTQWSKNMKHLTPSTLTQIDYNEKEKGAITQSPSLCLTRSHI 1380
Qy 1400 PQANRSPPIAKVSPFPIRPIYLRVLFDQNSHLPAAVSRKDKSGVQESSHFPQGAKK 1459
Db 1381 PQANRSPPIAKVSPFPIRPIYLRVLFDQNSHLPAAVSRKDKSGVQESSHFPQGAKK 1440
Qy 1460 NNLISLAIITLEMTGQREVGSLGTATNSVYKVENVTLPKPDLPKTSKVELLPKVEI 1519
Db 1441 NNLISLAIITLEMTGQREVGSLGTATNSVYKVENVTLPKPDLPKTSKVELLPKVEI 1500
Qy 1520 YQKOLFPETISNGSPCHLDLVEGSLQGTGEGAIKWEANRPGKVPFLVATBESSAKTPSK 1579
Db 1501 YQKOLFPETISNGSPCHLDLVEGSLQGTGEGAIKWEANRPGKVPFLVATBESSAKTPSK 1560
Qy 1580 LLDPLAWDNHYGTQIPKEBWSQESPEKTAFFKXDTILSLNACESNHAIAINSGQNK 1639
Db 1561 LLDPLAWDNHYGTQIPKEBWSQESPEKTAFFKXDTILSLNACESNHAIAINSGQNK 1620
Qy 1640 EIEVTWAKQGTRELCSQNPVLPKHQBREITRTTQSQDEEIDYDITISVENMKXDFDIY 1699
Db 1621 EIEVTWAKQGTRELCSQNPVLPKHQBREITRTTQSQDEEIDYDITISVENMKXDFDIY 1680
Qy 1700 DEDENQSPRSFQKTRHYFIAAVERLWBYGMSSSPHVLNRNRAQSGSVQPKVQFQETD 1759
Db 1681 DEDENQSPRSFQKTRHYFIAAVERLWBYGMSSSPHVLNRNRAQSGSVQPKVQFQETD 1740
Qy 1760 GSFTOPLYRGEALNEHLGLGPYIRAEVEDNIMVTFRNQASRPYSFYSLISYEEDQROGA 1819
Db 1741 GSFTOPLYRGEALNEHLGLGPYIRAEVEDNIMVTFRNQASRPYSFYSLISYEEDQROGA 1800
Qy 1820 BRKKNVKNENETKTYFWKQVHHMAPTKQBFDCAWAYFSDVLEKDXVHSGLGLPLLVCHT 1879
Db 1801 BRKKNVKNENETKTYFWKQVHHMAPTKQBFDCAWAYFSDVLEKDXVHSGLGLPLLVCHT 1860
Qy 1880 NTLNPAHGRQVTVQBFALFTTIFDETKSWYFTENNERNCRAPCNIQMEDPTFKENYRFA 1939
Db 1861 NTLNPAHGRQVTVQBFALFTTIFDETKSWYFTENNERNCRAPCNIQMEDPTFKENYRFA 1920

Qy 1940 INGVIYMDTLPGLVMAQDQRIRWYLLSMGSNENIHSIHFSGHVFTVRKKEEYKALYNLYP 1999
Db 1921 INGVIYMDTLPGLVMAQDQRIRWYLLSMGSNENIHSIHFSGHVFTVRKKEEYKALYNLYP 1980
Qy 2000 GVFTTEVMLPSKAGIWRVECLIGHLHAGMSTLFLVYSNKCQTPLGMAASGHIRDFQTAS 2059
Db 1981 GVFTTEVMLPSKAGIWRVECLIGHLHAGMSTLFLVYSNKCQTPLGMAASGHIRDFQTAS 2040
Qy 2060 GOYQOWAPKLARLHYSGSINAWSTKEPFWTKVLLAPMIHGIKTGACRKPSSLYISO 2119
Db 2041 GOYQOWAPKLARLHYSGSINAWSTKEPFWTKVLLAPMIHGIKTGACRKPSSLYISO 2100
Qy 2120 FLIYMSLDGKKWQYRGNSTGTLVFFGNVDSSGIKNIENPPIIARIYIRLEPHYSIRS 2179
Db 2101 FLIYMSLDGKKWQYRGNSTGTLVFFGNVDSSGIKNIENPPIIARIYIRLEPHYSIRS 2160
Qy 2180 TLRMELMGCDLNSCMLPGMBESKAISDAQITASSYFTNMFATWSPSKARLHLQGRSNAWR 2239
Db 2161 TLRMELMGCDLNSCMLPGMBESKAISDAQITASSYFTNMFATWSPSKARLHLQGRSNAWR 2220
Qy 2240 PQVANNPKEMLOVDFOKTKVKTGVTTOGVKSLTTSWYKKEPLISSQDGHOWTLFPQNGKV 2299
Db 2221 PQVANNPKEMLOVDFOKTKVKTGVTTOGVKSLTTSWYKKEPLISSQDGHOWTLFPQNGKV 2280
Qy 2300 KVFQGNQDSFTPVVNSLDDPPLLTRYLRIHPQSWVHQIALRMEVLGCEAQDIY 2351
Db 2281 KVFQGNQDSFTPVVNSLDDPPLLTRYLRIHPQSWVHQIALRMEVLGCEAQDIY 2332

RESULT 10

US-10-445-235-2
; Sequence 2, Application US/10445235
; Publication No. US20040005670A1
; GENERAL INFORMATION:
; APPLICANT: Katherine A. High
; APPLICANT: Rodney M. Canlire
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE
; FILE REFERENCE: CHOP0176
; CURRENT APPLICATION NUMBER: US/10/445,235
; CURRENT FILING DATE: 2003-05-22
; PRIOR APPLICATION NUMBER: 60/382,486
; PRIOR FILING DATE: 2002-05-22
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 2332
; TYPE: PRT
; ORGANISM: homo sapiens
US-10-445-235-2

Query Match 98.9%; Score 12282; DB 15; Length 2332;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 2327; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Qy 20 ATRRYLGAVELSDYMQSDLGELPVDARFPPRPVKSPFNTSVYKTLFVEFTDHLFN 79
Db 1 ATRRYLGAVELSDYMQSDLGELPVDARFPPRPVKSPFNTSVYKTLFVEFTDHLFN 60
Qy 80 IAKRPPPMWGLGTTIQAEVDTVTITLKNWASHPVSLHAGVSVYKASGAEYDDQTSQ 139
Db 61 IAKRPPPMWGLGTTIQAEVDTVTVTITLKNWASHPVSLHAGVSVYKASGAEYDDQTSQ 120
Qy 140 REKEDDKVFGSGSHYVYQVLKENGPMASDPLCTIYSLVSHVDIVKDLNSGLIGALLVCR 199
Db 121 REKEDDKVFGSGSHYVYQVLKENGPMASDPLCTIYSLVSHVDIVKDLNSGLIGALLVCR 180
Qy 200 EGS LAKEKTQTLHFIILLFAVDFDECKSWHSETKNSLMQDRDAASARAWPKHVTNGYVNR 259
Db 181 EGS LAKEKTQTLHFIILLFAVDFDECKSWHSETKNSLMQDRDAASARAWPKHVTNGYVNR 240
Qy 260 SLPLGICHRKSVVHWIYGMGTTTPEVHSIFLEGHTFLVNRHQASLEISPTTFTTAQTLL 319

Db 241 SLPGLIGCHRKGVYWEHVGWGTTPVHVSIFLBGHTPLVNRHQASLEISPIITFLTAQTL 300
QY 320 MDLGGFLLSCHTSSHOHDGWEAYVKVDSCEBPQLIMKNNEAABDYDDLTDSMDVVRP 379
Db 301 MDLGGFLLSCHTSSHOHDGWEAYVKVDSCEBPQLIMKNNEAABDYDDLTDSMDVVRP 360
QY 380 DDNSPFSFIQIRSAVKHFKPTWVHYIAAEEEDWDVAPLVLAPDDRSYKSOYLNNPQIRIG 439
Db 361 DDNSPFSFIQIRSAVKHFKPTWVHYIAAEEEDWDVAPLVLAPDDRSYKSOYLNNPQIRIG 420
QY 440 RYKXKVRMAYTDETFKTEAIOHSGILGPLLYGEVGTLLIIIPKQASRPYNIYPHGI 499
Db 421 RYKXKVRMAYTDETFKTEAIOHSGILGPLLYGEVGTLLIIIPKQASRPYNIYPHGI 480
QY 500 TDVRPLYSRSLPKGVKHLXDFPILPGEIEKYKWTIVVEDGPKSPRCLTRYSSPVMME 559
Db 481 TDVRPLYSRSLPKGVKHLXDFPILPGEIEKYKWTIVVEDGPKSPRCLTRYSSPVMME 540
QY 560 RDLASGLIGPLIICYKESVDQGNQIMSDKXNVILFVSFDENRSWYLTENIORFLPNPAG 619
Db 541 RDLASGLIGPLIICYKESVDQGNQIMSDKXNVILFVSFDENRSWYLTENIORFLPNPAG 600
QY 620 VQLEDPFOASIMHSINGYFDSLSOLSVCLHEVAVWYILSICAQDPLSVFSGYTFKH 679
Db 601 VQLEDPFOASIMHSINGYFDSLSOLSVCLHEVAVWYILSICAQDPLSVFSGYTFKH 660
QY 680 KMVYEDTLTFPFSGETVPMENPCLWILGCHNSDFRNKGTALLKVSSCDKNIGDYVE 739
Db 661 KMVYEDTLTFPFSGETVPMENPCLWILGCHNSDFRNKGTALLKVSSCDKNIGDYVE 720
QY 740 DSYEDISAYLLSKNAIEPRFSQNSRHPSTRQKQFNATTIPENDIEKTDPMFAHRTMP 799
Db 721 DSYEDISAYLLSKNAIEPRFSQNSRHPSTRQKQFNATTIPENDIEKTDPMFAHRTMP 780
QY 800 KIONVSSDMLLKQSPFPHGLSLSDLQAKYETPSDDPSGCAIDSNNSLSEMTFRPQ 859
Db 781 KIONVSSDMLLKQSPFPHGLSLSDLQAKYETPSDDPSGCAIDSNNSLSEMTFRPQ 840
QY 860 LHSQDMVFTPSGQLRNEKLGITTAATELKKLDFKVSSTNNLITSTPSNLAAGTUN 919
Db 841 LHSQDMVFTPSGQLRNEKLGITTAATELKKLDFKVSSTNNLITSTPSNLAAGTUN 900
QY 920 TSSLOPPMPHYDOLDTTLFGKXESPLTESGGPLSSEENNSDKLLESGLMNSQESSW 979
Db 901 TSSLOPPMPHYDOLDTTLFGKXESPLTESGGPLSSEENNSDKLLESGLMNSQESSW 960
QY 980 GKNVSTESGRIPKGRAGPALLTKDNALFKVSI SLLKTNKTSNNSATNRKTHIDGPSL 1039
Db 961 GKNVSTESGRIPKGRAGPALLTKDNALFKVSI SLLKTNKTSNNSATNRKTHIDGPSL 1020
QY 1040 LIENSPPSVNONTLESDETEPKVTPLIHDMMLMDKNATALRLNHSNKTSSKNXEMVQOK 1099
Db 1021 LIENSPPSVNONTLESDETEPKVTPLIHDMMLMDKNATALRLNHSNKTSSKNXEMVQOK 1080
QY 1100 KEGPIPPDQNDPMSFPKMLFIPESARWQORTHGKNSLNSGQSPKQIVSLGPEKSVEG 1159
Db 1081 KEGPIPPDQNDPMSFPKMLFIPESARWQORTHGKNSLNSGQSPKQIVSLGPEKSVEG 1140
QY 1160 QNPLSEKKNVYKCBFTXOVGLKENVPDSSRNPLFTNLDNLHENTHNOEKKIOBEIEK 1219
Db 1141 QNPLSEKKNVYKCBFTXOVGLKENVPDSSRNPLFTNLDNLHENTHNOEKKIOBEIEK 1200
QY 1220 KETLIOENVVLPOIHVVTGTQKFMKNLFLISTRQNVESGDYAGYAPVLODFRSLNDSTNR 1279
Db 1201 KETLIOENVVLPOIHVVTGTQKFMKNLFLISTRQNVESGDYAGYAPVLODFRSLNDSTNR 1260
QY 1280 TKXHTAHFKKGBEENLEGLNQTKQIVKQYACTTRISPTNSQONFVTORSKALKQFRL 1339
Db 1261 TKXHTAHFKKGBEENLEGLNQTKQIVKQYACTTRISPTNSQONFVTORSKALKQFRL 1320
QY 1340 PLEETELEXRIIVDDTSTOWSKXMKHLTSTLTQIDYNEKEKGAITQSPSLCPLRSHSI 1399
Db 1321 PLEETELEXRIIVDDTSTOWSKXMKHLTSTLTQIDYNEKEKGAITQSPSLCPLRSHSI 1380

QY 1400 PQANBSPPIAKVSPFPIRPIYLTRVLFOONSSHLPAASYKKDQSGYQESSHFLQGAKK 1459
Db 1381 PQANBSPPIAKVSPFPIRPIYLTRVLFOONSSHLPAASYKKDQSGYQESSHFLQGAKK 1440
QY 1460 NNLSIAIILTEMTGQREVGSIGTSATNSVTYKVKVENTVLPKPDLFKTSKGVKVELLPKVHI 1519
Db 1441 NNLSIAIILTEMTGQREVGSIGTSATNSVTYKVKVENTVLPKPDLFKTSKGVKVELLPKVHI 1500
QY 1520 YOKDLFPFETNNGSPGHLDLVEGSLQOTEGAIAKWEANRPGKVPFLRVAIETESSAKTFSK 1579
Db 1501 YOKDLFPFETNNGSPGHLDLVEGSLQOTEGAIAKWEANRPGKVPFLRVAIETESSAKTFSK 1560
QY 1580 ILDPLAWNHYGTQIPKBEWKSQKSPKTAFAKXDTILSLNACESNHAIAAINEGQNK 1639
Db 1561 ILDPLAWNHYGTQIPKBEWKSQKSPKTAFAKXDTILSLNACESNHAIAAINEGQNK 1620
QY 1640 BIEVTWAKQGRTERLCSQNPVLRKHOREIITRTILSQDQEEIDYDDTISVEMKKEDFDIY 1699
Db 1621 BIEVTWAKQGRTERLCSQNPVLRKHOREIITRTILSQDQEEIDYDDTISVEMKKEDFDIY 1680
QY 1700 DBDENQSPRSFQKTRHYFIAAVERLWDYGMSSSPHVLNRQAQSGVFPQFKVVPQEEFTD 1759
Db 1681 DBDENQSPRSFQKTRHYFIAAVERLWDYGMSSSPHVLNRQAQSGVFPQFKVVPQEEFTD 1740
QY 1760 GSFTOPLVRGELNEHLGLLGPYIRAEVDNIMVTFRNOASRPYSYSSILSYEEDOROGA 1819
Db 1741 GSFTOPLVRGELNEHLGLLGPYIRAEVDNIMVTFRNOASRPYSYSSILSYEEDOROGA 1800
QY 1820 EPRKNFVKPNETKTYFWKVQHMAPTKDFCKAWAYFSDVDLEKDVHSGLIGPLLVCHT 1879
Db 1801 EPRKNFVKPNETKTYFWKVQHMAPTKDFCKAWAYFSDVDLEKDVHSGLIGPLLVCHT 1860
QY 1880 NTLNPAHGRQVTVQBFALPFTIFDBTKSWYFTENNERNCRAPCNQMEDPTFKENYRPHA 1939
Db 1861 NTLNPAHGRQVTVQBFALPFTIFDBTKSWYFTENNERNCRAPCNQMEDPTFKENYRPHA 1920
QY 1940 INGYIMDTLPLGLVMAQDQRIKWLILMGSNENIHSIHFSGHVFTVRKKEEKYKALYNLYP 1999
Db 1921 INGYIMDTLPLGLVMAQDQRIKWLILMGSNENIHSIHFSGHVFTVRKKEEKYKALYNLYP 1980
QY 2000 GVFEVTEMLPSKAGIWRVECCJGHEHLAGMSTPLVYSNKQCTP-GMASGHIRDPQITAS 2059
Db 1981 GVFEVTEMLPSKAGIWRVECCJGHEHLAGMSTPLVYSNKQCTPGLGMASGHIRDPQITAS 2040
QY 2060 GOYQOWAPKLARLHVSIGSINAWSTKEPFSWIKVDLLAPMI IHGIKTQAGARQKPSLIYSQ 2119
Db 2041 GOYQOWAPKLARLHVSIGSINAWSTKEPFSWIKVDLLAPMI IHGIKTQAGARQKPSLIYSQ 2100
QY 2120 FTIMYSLDGKKWQTVRGNSGTGLMVFFGNVDSSGKKNIFNPPIIARVIRLHETHYSIRS 2179
Db 2101 FTIMYSLDGKKWQTVRGNSGTGLMVFFGNVDSSGKKNIFNPPIIARVIRLHETHYSIRS 2160
QY 2180 TLRMELMGCDLNSCNPGLMESKAIISDAQITASSYFTNNFATWSFSKARLHLQGSNAWR 2239
Db 2161 TLRMELMGCDLNSCNPGLMESKAIISDAQITASSYFTNNFATWSFSKARLHLQGSNAWR 2220
QY 2240 PQVNNPKEWLQVDPQKTKVGTITQGVKSLTTSYVVEFLISSQDGHOWTLFPQNGKV 2299
Db 2221 PQVNNPKEWLQVDPQKTKVGTITQGVKSLTTSYVVEFLISSQDGHOWTLFPQNGKV 2280
QY 2300 KVFQNGQDSFPVNSLDPPLLTLYRIRTHPSQVHQAIALRMEVLGCEAQQDLY 2351
Db 2281 KVFQNGQDSFPVNSLDPPLLTLYRIRTHPSQVHQAIALRMEVLGCEAQQDLY 2332

RESULT 11

US-10-239-498A-2
; Sequence 2, Application US/10239498A
; Publication No. US2004002333A1
; GENERAL INFORMATION:
; APPLICANT: Hauser, Charlotte
; APPLICANT: Horster, Andrea

APPLICANT: Schroder, Carola
APPLICANT: Lehner, Michael
TITLE OF INVENTION: Production of Recombinant Blood Clotting Factors in
TITLE OF INVENTION: Human Cell Lines
FILE REFERENCE: 80577.0001
CURRENT APPLICATION NUMBER: US/10/239,458A
CURRENT FILING DATE: 2003-07-08
PRIOR FILING DATE: 2001-03-21
NUMBER OF SEQ ID NOS: 17
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 2
LENGTH: 2332
TYPE: PRT
ORGANISM: Homo sapiens
US-10-239-458A-2

Query Match 98.9%; Score 12282; DB 16; Length 2332;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 2327; Conservative 2; Mismatches 3; Indels 0; Gaps 0;
QY 20 ATRYYILGAVELSDWYMSDGLGELPVDARPPRPVKSPFFNTSVYVYKTLFVETDHLFN 79
DB 1 ATRYYILGAVELSDWYMSDGLGELPVDARPPRPVKSPFFNTSVYVYKTLFVETDHLFN 60
QY 80 IAKRPPPMWGLGPTIQAEVYDVTIVILKNVASHPVSLHAYGVSYWKASGAEDDQTSQ 139
DB 61 IAKRPPPMWGLGPTIQAEVYDVTIVILKNVASHPVSLHAYGVSYWKASGAEDDQTSQ 120
QY 140 REKEDDKVPPGSGTYVYVQVILKENGPMASDPLCTYSVLSHVDVLKOLNSGLIGALLVCR 199
DB 121 REKEDDKVPPGSGTYVYVQVILKENGPMASDPLCTYSVLSHVDVLKOLNSGLIGALLVCR 180
QY 200 EGS LAKEKTOFLHFKLILLFAVFDGKSWHSTKNSLMQDRDAASARAWPKMHTVNGYNR 259
DB 181 EGS LAKEKTOFLHFKLILLFAVFDGKSWHSTKNSLMQDRDAASARAWPKMHTVNGYNR 240
QY 260 SLPLGIGHKRSVYVHVGMTTPEVHSIFLEIGHTFLVRNHRQASLEISPTFLTAQTLL 319
DB 241 SLPLGIGHKRSVYVHVGMTTPEVHSIFLEIGHTFLVRNHRQASLEISPTFLTAQTLL 300
QY 320 MDLQGFLLSCHISFQHDGMAVYKVDSCPEEPQILMKNNEEADYDDDLTDSMDVVRP 379
DB 301 MDLQGFLLSCHISFQHDGMAVYKVDSCPEEPQILMKNNEEADYDDDLTDSMDVVRP 360
QY 380 DDNSPSFIQIRSVAKKHPXTWVHYIAAEEDWDYAPVLAPDDRYSYKSOYLNNGPQRIG 439
DB 361 DDNSPSFIQIRSVAKKHPXTWVHYIAAEEDWDYAPVLAPDDRYSYKSOYLNNGPQRIG 420
QY 440 RYKAKVRFMAYTDETFKTRZAIQHSGLILGPLYGEVGDITLLIIFKQASRPYNIYPHGI 499
DB 421 RYKAKVRFMAYTDETFKTRZAIQHSGLILGPLYGEVGDITLLIIFKQASRPYNIYPHGI 480
QY 500 TDVREPLYSRRLPKGVKHLKDPILPGEIFKVTWVEDGPTKSDPRCLTRYVSSFVYVME 559
DB 481 TDVREPLYSRRLPKGVKHLKDPILPGEIFKVTWVEDGPTKSDPRCLTRYVSSFVYVME 540
QY 560 RDLASGLIGPLLIICYKESVDQGNQIMSDKRNVLFSVFDENRSWYLTENIQRLPNPAG 619
DB 541 RDLASGLIGPLLIICYKESVDQGNQIMSDKRNVLFSVFDENRSWYLTENIQRLPNPAG 600
QY 620 VQLEDPEQASNIHMSINGVYFDSQLQSVCLHEVAYWYILSIGAQDTFLSVFFSGYTFKH 679
DB 601 VQLEDPEQASNIHMSINGVYFDSQLQSVCLHEVAYWYILSIGAQDTFLSVFFSGYTFKH 660
QY 680 KMYVEDTLTLPPFSGEIVFMSMENPGLWILGCHNSDFRNRCWTALLKVVSCDKNTGDIYE 739
DB 561 KMYVEDTLTLPPFSGEIVFMSMENPGLWILGCHNSDFRNRCWTALLKVVSCDKNTGDIYE 720
QY 740 DSYEDIAYLLSKNAIEPRFSQNSRHPSTRQGFNATTIPENDIEKTPWFFAHRTMP 799
DB 721 DSYEDIAYLLSKNAIEPRFSQNSRHPSTRQGFNATTIPENDIEKTPWFFAHRTMP 780

QY 800 KIQVSSSDLLMLLRQSPTHGLISLSDLOEAXYETFSDDPSGAIDSNNSISEMTHFRPQ 859
DB 781 KIQVSSSDLLMLLRQSPTHGLISLSDLOEAXYETFSDDPSGAIDSNNSISEMTHFRPQ 840
QY 860 LHSQDMVFTPESGLOLRINEKLGITTAATELKKLDLKFVSSSTNNLISTIPSDNLAAGTDN 919
DB 841 LHSQDMVFTPESGLOLRINEKLGITTAATELKKLDLKFVSSSTNNLISTIPSDNLAAGTDN 900
QY 920 TSSLGPPSMVHYDSQDITLFGKKSPLTSGGPLSLSEENNDKSLLESGLMNSQSSW 979
DB 901 TSSLGPPSMVHYDSQDITLFGKKSPLTSGGPLSLSEENNDKSLLESGLMNSQSSW 960
QY 980 GKNVSSTESGRLFKGKAHGPALLITKONALPKVSIILKTKNTKSNNSATNKTTHIDGSL 1039
DB 961 GKNVSSTESGRLFKGKAHGPALLITKONALPKVSIILKTKNTKSNNSATNKTTHIDGSL 1020
QY 1040 LIENSPPVQWNLSDTEFKKVTFLIHDRMLMDKNATALRLNHSNKTTSKMEMVQOK 1099
DB 1021 LIENSPPVQWNLSDTEFKKVTFLIHDRMLMDKNATALRLNHSNKTTSKMEMVQOK 1080
QY 1100 KEGPIPPDAQNPDMSFKMLFLPESARWIQETHGKNSINSCQGPSKOLVSLGPEKSVEG 1159
DB 1081 KEGPIPPDAQNPDMSFKMLFLPESARWIQETHGKNSINSCQGPSKOLVSLGPEKSVEG 1140
QY 1160 QNFLSEKXKVVVGKERTKDVGLKEMVFPSSRNFLTLNLDLHNNTHNQSKTKQEEIEK 1219
DB 1141 QNFLSEKXKVVVGKERTKDVGLKEMVFPSSRNFLTLNLDLHNNTHNQSKTKQEEIEK 1200
QY 1220 KETLIQENVVLPOIHTVGTGTFKMNFLFLSTRQNVESYDGAYAPVLQDPRSLDSTNR 1279
DB 1201 KETLIQENVVLPOIHTVGTGTFKMNFLFLSTRQNVESYDGAYAPVLQDPRSLDSTNR 1260
QY 1280 TKKHTAFHSKGBEENLEGIGNOTKOIYEVYACVTRISPNTSQONFVTRSKRALKQFRL 1339
DB 1261 TKKHTAFHSKGBEENLEGIGNOTKOIYEVYACVTRISPNTSQONFVTRSKRALKQFRL 1320
QY 1340 PLEETELEKXIIIVDDTSTQWSKMKHLTPSTLTQIDYNEKEKGAITOSPISDCLTRSHSI 1399
DB 1321 PLEETELEKXIIIVDDTSTQWSKMKHLTPSTLTQIDYNEKEKGAITOSPISDCLTRSHSI 1380
QY 1400 POANRSLPITAKVSSPFSIRPIYITRVLFQDNSSHLPAASVYRKDQGVQSSSHFLQAKK 1459
DB 1381 POANRSLPITAKVSSPFSIRPIYITRVLFQDNSSHLPAASVYRKDQGVQSSSHFLQAKK 1440
QY 1460 NNLSLAILTLTLEMTQDQREVSLGTSATNSVYKVENTVILPKDLPKTSKGVLLPKVHI 1519
DB 1441 NNLSLAILTLTLEMTQDQREVSLGTSATNSVYKVENTVILPKDLPKTSKGVLLPKVHI 1500
QY 1520 YQKDLPTETSNSSPGHLDLVEGSLLOGTEGAIKWNEANRPGKVPFLRVATESSAKTPSK 1579
DB 1501 YQKDLPTETSNSSPGHLDLVEGSLLOGTEGAIKWNEANRPGKVPFLRVATESSAKTPSK 1560
QY 1580 LLDPLAWNHYGTQIPKEEWKSOEKSPKTAFFKKDTILSLNACESHAAIAANEGONKP 1639
DB 1561 LLDPLAWNHYGTQIPKEEWKSOEKSPKTAFFKKDTILSLNACESHAAIAANEGONKP 1620
QY 1640 EIEVTWAKQRTERLCSQNPVILKHOEITRTTLTQSDQDEIDYDDTISVEMKKEDEDIY 1699
DB 1621 EIEVTWAKQRTERLCSQNPVILKHOEITRTTLTQSDQDEIDYDDTISVEMKKEDEDIY 1680
QY 1700 DEBENQSPRSFQKTRHYFTAAVERLWDYGMSSPHVLRNRAQSGSVPOPKVVFQFTD 1759
DB 1681 DEBENQSPRSFQKTRHYFTAAVERLWDYGMSSPHVLRNRAQSGSVPOPKVVFQFTD 1740
QY 1760 GSTQPLRYGELNEHLGLGPIYRAEVEDNIMVTFNQASRPYSFYSSLSIYSEDORQGA 1819
DB 1741 GSTQPLRYGELNEHLGLGPIYRAEVEDNIMVTFNQASRPYSFYSSLSIYSEDORQGA 1800
QY 1820 EPRKNTFKVNETKTYFWKVOHMAIPKDBDFCKAWAFSDVDLEKOVHSLIGPLLVCHT 1879
DB 1801 EPRKNTFKVNETKTYFWKVOHMAIPKDBDFCKAWAFSDVDLEKOVHSLIGPLLVCHT 1860
QY 1880 NTLNPAHQROVTYOEALPFTTIFDETCKSWYFTENMBENCRAPCNIQMEDPTFKENYRFA 1939

1861 NTLNPAHQRTVQBFALFFIFDETKSWFTENMERNCRAPNTQMEDPTPKENYRTHA 1920
1940 INGVIMDTLPGVMAQDQIRIKVLLSMGSENIHSHFSGHVFTRKKEEYKMAVLYNP 1999
1921 INGVIMDTLPGVMAQDQIRIKVLLSMGSENIHSHFSGHVFTRKKEEYKMAVLYNP 1980
2000 GYFTEVEMLPKAGIWRVECLIGEHLHAGMSTFLVYGNKQOTPLGMASGHIRDFQITAS 2059
1981 GYFTEVEMLPKAGIWRVECLIGEHLHAGMSTFLVYGNKQOTPLGMASGHIRDFQITAS 2040
2060 GYQGWAPKARLHVSIGSINAWSTKEPWSIKVDLLAPMIHGIKTOGARQKPSLYISQ 2119
2041 GYQGWAPKARLHVSIGSINAWSTKEPWSIKVDLLAPMIHGIKTOGARQKPSLYISQ 2100
2120 FIIMYSLQKKWQYRGNSGTGLMVFPGVNDSSGKHNIEPPIIARVIRLHPHYSIRS 2179
2101 FIIMYSLQKKWQYRGNSGTGLMVFPGVNDSSGKHNIEPPIIARVIRLHPHYSIRS 2160
2180 TLRMELMGCDLNSCSPMLGMSKALSDAQITASSYFTNWFATWSPKARLHLQGRSNAR 2239
2161 TLRMELMGCDLNSCSPMLGMSKALSDAQITASSYFTNWFATWSPKARLHLQGRSNAR 2220
2240 P0VANNPKWLOVDFOKTMKVTVGTTCQVKSLTSMYVKEPLISSQDGHQWTLFPQNGKV 2299
2221 P0VANNPKWLOVDFOKTMKVTVGTTCQVKSLTSMYVKEPLISSQDGHQWTLFPQNGKV 2280
2300 KVFQGNQDSFPTVNVSLDPPILLRLIRHPQSWHVOIALRMEVLGCEAODLY 2351
2281 KVFQGNQDSFPTVNVSLDPPILLRLIRHPQSWHVOIALRMEVLGCEAODLY 2332

RESULT 12
US-10-187-319-6

GENERAL INFORMATION:
APPLICANT: Lollar, John S.
TITLE OF INVENTION: Hybrid Human/Animal Factor VIII
NUMBER OF SEQUENCES: 40
CORRESPONDENCE ADDRESSES:
ADDRESS: Greenlee, Winner and Sullivan, P.C.
STREET: 537C Manhattan Circle Suite 201
CITY: Boulder
STATE: Colorado
COUNTRY: USA
ZIP: 80303

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/187,319
FILING DATE: 27-Aug-2002
CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 09/523,656
FILING DATE: 2000-03-10
APPLICATION NUMBER: US 09/037,601
FILING DATE: 1998-03-10
APPLICATION NUMBER: WO PCT/US97/11155
FILING DATE: 1997-06-26
APPLICATION NUMBER: US 08/670,707
FILING DATE: 1996-06-26

ATTORNEY/AGENT INFORMATION:
NAME: Greenlee, Lorraine L.
REGISTRATION NUMBER: 27,894
REFERENCE/DOCKET NUMBER: 75-95X
TELECOMMUNICATION INFORMATION:
TELEPHONE: 303/499-8080
TELEFAX: 303/499-8089
TITLE: Sequence of the Murine Factor VIII cDNA
JOURNAL: Genomics
VOLUME: 16

PAGES: 374-379
DATE: 1993
RELEVANT RESIDUES IN SEQ ID NO: 6: FROM 1 TO 2319
SEQUENCE DESCRIPTION: SEQ ID NO: 6:
US-10-187-319-6

Query Match 71.0%; Score 8820; DB 14; Length 2319;
Best Local Similarity 72.4%; Pred. No. 0;
Matches 1711; Conservative 232; Mismatches 365; Indels 54; Gaps 22;

QY 1 MQIELSTCFIFCLLRPCFSATRRYYLGAVELSWYMOQSD-LGELPVDARPPRPVKSPFF 59
Db 1 MQIALFACFFLSLNFECSSAIRRYLGAVELSWNYIQSDLLSVLHSDSRFLPRMSTSPFF 60
QY 60 NTSVYVYKKTAFVBTDLHFNIAKRPWGMGLLGTIOAEVYDVTVITLKNMASHPVSLHA 119
Db 61 NTSVYVYKKTAFVBTDLHFNIAKRPWGMGLLGTIWTVEHDTVITLKNMASHPVSLHA 120
QY 120 VGSYWKASGAEYDDDTQSREKEDCKVFPQGSSTVYVQVLKENGPMASDPLCLTYSYLS 179
Db 121 VGSYWKASGAEYDDDTQSREKEDCKVFPQGSSTVYVQVLKENGPMASDPLCLTYSYMS 180
QY 180 HVDLVKDLNSGLIGALLVCRGSLAKTQTLHFKFILLFAVDEGKSWHSTKNSLMQDR 239
Db 181 HVDLVKDLNSGLIGALLVCRGSLAKTQTLHFKFILLFAVDEGKSWHSTKNSLMQDR 240
QY 240 DAASARAWPKXHTVNGVYVNSLPLGLIGCHRKSVYVHVGTTPEVHSIFLEGHTELVRN 299
Db 241 DAASARAWPKXHTVNGVYVNSLPLGLIGCHRKSVYVHVGTTPEVHSIFLEGHTELVRN 300
QY 300 HRQASLEISPTITLTAQTLMLDGLQFLLSCHSISSHQHDGMEAYVYKVDSCPEPOLIMK-N 358
Db 301 HRQASLEISPTITLTAQTLMLDGLQFLLSCHSISSHQHDGMEAYVYKVDSCPEPOLIMK-N 360
QY 359 NEBAEDDDDLTDSEMDVVRPDDNSPSFTQIRVAKKHPTWVHYIAAREEDWDYAPLV 418
Db 361 NEEMEDYDDDLTDSEMDVVRPDDNSPSFTQIRVAKKHPTWVHYIAAREEDWDYAPLV 418
QY 419 LAPDRSVKSOYLANGPQIRGRKVKYRFMAYTDETFKTRAIQHEGSLGLPLLYGEVD 478
Db 419 PTDNGSYKSOYLANGPQIRGRKVKYRFMAYTDETFKTRAIQHEGSLGLPLLYGEVD 478
QY 479 TLLIIFKNQASRPYNIYPHGITDVRPLYSRRLPKGVKHLKDFPLPGEIFKYKWTIVED 538
Db 479 TLLIIFKNQASRPYNIYPHGITDVRPLYSRRLPKGVKHLKDFPLPGEIFKYKWTIVED 538
QY 539 GPTKSDRCLTRYSSFWNVKRDLAGSLIGPLLI CYKESVDQKGNQIMSKRNVLISVF 598
Db 539 GPTKSDRCLTRYSSFWNVKRDLAGSLIGPLLI CYKESVDQKGNQIMSKRNVLISVF 598
QY 599 DENRSWYLTENIQRFNPNAGVQLDEPFOASNTMHSINGVYVFDLSQLSVCLHEVAYWYI 658
Db 599 DENRSWYLTENIQRFNPNAGVQLDEPFOASNTMHSINGVYVFDLSQLSVCLHEVAYWYI 658
QY 659 LSIQAQTDVLSVFPSSGYTFKHMYIEDTLTLFPSSGETVFMENPGLWILGCNSDFRN 718
Db 659 LSIQAQTDVLSVFPSSGYTFKHMYIEDTLTLFPSSGETVFMENPGLWILGCNSDFRN 718
QY 719 RGMNTALLKVSQCDNTGDDYEDSDYEDISAVLLSKNNAIEPRSPQNSHSPSTROKQFAT 778
Db 719 RGMNTALLKVSQCDNTGDDYEDSDYEDISAVLLSKNNAIEPRSPQNSHSPSTROKQFAT 778
QY 779 TIPENDIEKTDPMFAHRTMPKIQNVSSSDLLMLLROS-PTEHGLSLSDLOEAKYETFS 837
Db 779 TIPENDIEKTDPMFAHRTMPKIQNVSSSDLLMLLROS-PTEHGLSLSDLOEAKYETFS 837
QY 838 DPSPGALDSSNLSJENMTHFRPQLHSGDMVFTPESGQLRLNEXLGTATAELKLLDFKV 897
Db 838 DPSPGALDSSNLSJENMTHFRPQLHSGDMVFTPESGQLRLNEXLGTATAELKLLDFKV 897
QY 898 DHPENADSDNEGPKVTQLRPESSHSEKIVFTPOGQLRNSKSLTTEVKKWKLGLQV 956
Db 898 DHPENADSDNEGPKVTQLRPESSHSEKIVFTPOGQLRNSKSLTTEVKKWKLGLQV 956
QY 956 SSTSNLLI-STIPSDNLAAGTDNTSSIGPSPMPVHYDSQDITLPGKSSPLTESSGGLS 956
Db 956 SSTSNLLI-STIPSDNLAAGTDNTSSIGPSPMPVHYDSQDITLPGKSSPLTESSGGLS 956

Qy 957 LSEBNDKLLBSGIMNSQESSWGNVSSSTSCRLFKCKRAHFGPALLTKONALFKVUSISL 1016
Db 959 ASEENSDNLDSTLMYSQESPLRNDLUSIENCLREKRPFHGIALTKONTLPKDNVSL 1018
Qy 1017 LKTKNTSNNATNKTIDGSLSIENSPPVQN--ILESDFEFKKTPLIHDRMLMDKNA 1075
Db 1019 MKTKNTYHSTNEKLHTSEFT-SIENSTTDLQDAILKVNSEIOEVTALIHGDTLLGKNS 1077
Qy 1076 TALBNHMSNTTSGKNMVMQCKEGBIPDDAQNDKSPFKMLPLPESARWIORHCKN 1135
Db 1078 TYLSLNHMLNRTTGNKNDI ZHRKOEDIPQDEENTIIPFFSKMLFLSESSWPKTKNGN 1137
Qy 1136 SLNSGQSPKCLVSLGPEKSEVGNFSEKKNVVGKGETKOVGLKEMVPPSSRNLF 1195
Db 1138 SLNBEQESPKQLVYLMFKCKVKNQSFSEKKNVTVQDGTGKNGIKGLKDMAPPNMSIFL 1197
Qy 1196 TNLONLHNTHNQFKIQEBIEKETHIQENVLFPQIHTVTGTRNFMKNLFLASTQNV 1255
Db 1198 TLLSNVHENGHNQEKIQEBIE-KEALIEKVLPQVHEATGSKNFKLDILIGTRQNI 1256
Qy 1256 EGSYDGAVALQDFRSLNDSTNRKKTATPSK--KGEENLEGLNQTQVIVKCYACT 1313
Db 1257 --SLYEVHVPVLQNTITSNNSTNTVQIEMHFFFKRKDKETNSEGLVNKTREVMKNY--- 1311
Qy 1314 TRISNTSQNFVOTORSRAALKQRLPLEETELEKRI--VDDTSTOWSKNMKHLTPSTLTQ 1373
Db 1312 -----PSKNIITORSRAALKQRL-----STOWLKTIKNSOCIIKQ 1349
Qy 1374 IDYNEKEGALTQSPSLDCLTRSHSIPQANSPLPIAKVSPSPSIRPIYLRVLFDQNS 1433
Db 1350 IDHSEKMKFITKSLSDS-SVIKTTQNTSSDSHVKTSAFP---PIDLKSPQNKFS 1405
Qy 1434 HLPAAZY-----RKDSGVQBSHFLOGAKNNLSLAILEMTGQREVGSIGTSATNSV 1489
Db 1406 HVQASSYTFDEKTSRIGESNNFLKTKINNPSLAIPWNFIQGRFTSGKSTNSV 1465
Qy 1490 TYKVENTVLKPDLPKTSKGVKELLPKVHIYOKLPFTETNSGPHLDLVEGSLQOTE 1549
Db 1466 TYKRENTIFLPTUPESGKELLQPVQSIQEBILPTETSHGSPHNLNMEKVEFLQIK 1525
Qy 1550 GAIKWEANRFGKVPFLRVATESSAKTFSKLLDPLAWNHYQTQIPKEWKSQESPKPT 1609
Db 1526 GPTKWKAKRHE--SIGKTBSSKNTRSKLNHAWDYHVAQIPKDMWKSKEKSPRII 1583
Qy 1610 AFKDKDTILSNACSENEAIAINEGONKPEITVWAKQGRTERLCSQNPVYKRBQEI 1669
Db 1584 STKQEDTILSRPHGNSHISGA-NKQNPQRETWVGQGTQRTCSQIPPVYKRBQREL 1642
Qy 1670 TRTT--QSDEEDYDDTISVENMKKEDFDIYDBDENQSPRSFQKTRHYPIAAVERLWDY 1729
Db 1643 --SAFQSEQEAITYDAITIE-TIEDFDIYSEDIKQGRPSFQKTRHYPIAAVERLWDY 1699
Qy 1730 MSSSPHVLNRAQSSVPOFKKVVQBEFTDGSFTOPLRYGELNEHLGLLGPVIRAEVDN 1789
Db 1700 MST--HVLNRYQSDNVPOFKKVVQBEFTDGSFQPLRYGELNEHLGLLGPVIRAEVDN 1758
Qy 1790 INVTFRNOASRYSFYSSLSIYSSEDOQGAERKQFVKENETKVPKWQHHMAPTKDF 1849
Db 1759 INVTFRNOASRYSFYSSLSIYSSEDOQGAERKQFVKENETKVPKWQHHMAPTKDF 1817
Qy 1850 DCKAWAYFSDVLEKDVHSGSL-GPLIVCHTNTLNPAHGRQVTVQEFALFTTFIDETKSY 1909
Db 1818 DCKAWAYFSDVLEKDVHSGSLGIPULLICHANTLNPAHGRQVTVQEFALFTTFIDETKSY 1877
Qy 1910 FTENMERNACPCNIOMEDPTFKENTRYFALINGYINDTLPGLVMAQDQIRVYLLSMGN 1969
Db 1878 FTENVRNCKTPCNFOMEDPTLKENRYFALINGYINDTLPGLVMAQDQIRVYLLSMGN 1937
Qy 1970 ENIESHFSGHVFYVKKSEYKMALNLYPGVFETVEMLPKAGIWRVCLLGEHLHAGM 2029
Db 1938 ENIQSIHSGHVFYVKKSEYKMAVNLVPGVFETLXTPSRAGIWRVCLLGEHLQAGM 1997

Qy 2030 STLFLVYSNKOTPLMGASHIRDFQITASQYQGWAPKLARLHYSGSINAWSTKEPFSW 2089
Db 1998 STLFLVYSKCOQIPLVNASOSIRDFQITASGHYQGWAPNARLHYSGSINAWSTKEPFSW 2057
Qy 2090 IKVDLLAPMIHGIKTQAGARQSSLYISOFIIMYSIDGKKWQTYRGNSTGTLVWFFGNV 2149
Db 2058 IKVDLLAPMIHGIKTQAGARQSSLYISOFIIMYSIDGKKWLYSQGNSTGTLVWFFGNV 2117
Qy 2150 DSSGIKENIENPPIIARYIRLHPHYSIRSTLARMELMGCDLNSCMPLGMEKSAISDAQI 2209
Db 2118 DSSGIKNSNPNPIIARYIRLHPHYSIRSLARMELMGCDLNSCISPLGMEKSVISDTQI 2177
Qy 2210 TASSYFTNMFATWSPSKARLHLQGRSNARPNQVNNPKNWQVDFQKTMKVTVTQGVKS 2269
Db 2178 TASSYFTNMFATWSPSQARLHLQGRSNARPNQVNDPKWLQVDLQKTMKVTVTQGVKS 2237
Qy 2270 LLTSMYKPEFLISSODGHQWTLFPQKGVKVFQGNQDSFTPVVNSLDPLLTAYLRHP 2329
Db 2238 LFTSMYKPEFLISSODGHQWTLFYNGKVKVFQGNQDSSTPMNNSLDPLLTAYLRHP 2297
Qy 2330 QSWHJOIALRMEVLGCEAQQDLY 2351
Db 2298 QIWEHQ-ALRDEILGCEAQQY 2319
RESULT 13
US-10-131-510A-6
; Sequence 6, Application US/10131510A
; Publication No. US20030166536A1
; GENERAL INFORMATION:
; APPLICANT: Lollar, John S
; TITLE OF INVENTION: Modified Factor VIII
; FILE REFERENCE: 75-95J
; CURRENT APPLICATION NUMBER: US/10/131,510A
; CURRENT FILING DATE: 2002-07-10
; PRIOR APPLICATION NUMBER: U.S. 09/315,179
; PRIOR FILING DATE: 1999-05-20
; PRIOR APPLICATION NUMBER: U.S. 09/037,601
; PRIOR FILING DATE: 1998-03-10
; PRIOR APPLICATION NUMBER: U.S. 08/670,707
; PRIOR FILING DATE: 1996-06-26
; PRIOR APPLICATION NUMBER: PCI/US97/11155
; PRIOR FILING DATE: 1997-06-26
; PRIOR APPLICATION NUMBER: PCI/US94/13200
; PRIOR FILING DATE: 1994-11-15
; PRIOR APPLICATION NUMBER: U.S. 08/212,133
; PRIOR FILING DATE: 1994-03-11
; PRIOR APPLICATION NUMBER: U.S. 07/864,004
; PRIOR FILING DATE: 1992-04-07
; NUMBER OF SEQ ID NOS: 40
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 6
; LENGTH: 2319
; TYPE: PRT
; ORGANISM: Mus musculus
US-10-131-510A-6
Query Match 71.0%; Score 8820; DB 14; Length 2319;
Best Local Similarity 72.4%; Pred. No. 0;
Matches 1711; Conservative 232; Mismatches 365; Indels 54; Gaps 22;
Qy 1 MQEILSTCFFLCLAFCSATRRYYLGAVELSWDMQSD-LGELPVDARPPRPVPSPPF 59
Db 1 MQIALFAFFFLSLFNFCSSAIRYYLGAVELSNWYIQSDLLSVLHSDSRFLPRMSTSPFF 60
Qy 60 NTSVYVYKTLFVEPTDELFNIAKPPPMWGLLGPITQAEVYDTVITLKNMASHVSLHA 119
Db 61 NTSIMYKTKTFVEYKQDLFNIAKPPPMWGLLGPITWTEVHDVITLKNMASHVSLHA 120
Qy 120 VGVSYWKASEGAEYDDQTSQREKDDKVPFGSGSHYVWQVLKXENGPMSADPLCLTYSYLS 179
Db 121 VGVSYWKASEGDEYEDQTSQREKDDKVPFGSGSHYVWQVLKXENGPMSADPLCLTYSYMS 180

180 QY HVDLVKOLNSGLIGALLVCRSGLAKEKTQTLHKETILLFAVDECKSWHSETQNSLMQDR 239
181 Db HVDLVKOLNSGLIGALLVCRSGLAKEKTQTLHKETILLFAVDECKSWHSETQNSLMQDR 240
240 QY DAASARAPKMTTNGYVNRSLPGLIGCHRKQVYVHVMGTTPPEVHIFLEGHTFLVRN 299
241 Db DAASARAPKMTTNGYVNRSLPGLIGCHRKQVYVHVMGTTPPEVHIFLEGHTFLVRN 300
300 QY HRCASLETSPITFLTAOTLMDLGOFLLSCHSSHOCHDQMEAYVAVKDVCPPEPOLIMK-N 358
301 Db HRCASLETSPITFLTAOTLMDLGOFLLSCHSSHOCHDQMEAYVAVKDVCPPEPOLIMK-N 360
359 QY NDEAEYDDDLTSEYVVRPDDNSPFIQIRSAKHPKXTWVHYIAAEEDCDWYAPLV 418
361 Db NDEAEYDDDLTSEYVVRPDDNSPFIQIRSAKHPKXTWVHYIAAEEDCDWYAPLV 418
419 QY LAPDRSRYKQVNLNNGPQIRGKYKRYPMATVDETFKTREAIQHESGILGLLYGEVGD 478
419 Db LAPDRSRYKQVNLNNGPQIRGKYKRYPMATVDETFKTREAIQHESGILGLLYGEVGD 478
479 QY TLLIIFKQASRPYNIYPHGIIDVAPLYSRRLPKGVGLKDPPIIPGGEIFKYKWTWVED 538
479 Db TLLIIFKQASRPYNIYPHGIIDVAPLYSRRLPKGVGLKDPPIIPGGEIFKYKWTWVED 538
539 QY GPTKSDPCLTYYSFVVMERDLASGLIGPLLCYKESVDQKQIMSDKRNVLFSVF 598
539 Db GPTKSDPCLTYYSFVVMERDLASGLIGPLLCYKESVDQKQIMSDKRNVLFSVF 598
599 QY DENRSWYLTENORLPLNAGVQLDEPPEOASNIMHSINGYVDFSLQSVCHVAYMYI 658
599 Db DENRSWYLTENORLPLNAGVQLDEPPEOASNIMHSINGYVDFSLQSVCHVAYMYI 658
659 QY LSIQAQTDPLSFYFFGYTFKHMYVEDTLTLPFGSETVFMENPGLMILGCHNSDPEN 718
659 Db LSIQAQTDPLSFYFFGYTFKHMYVEDTLTLPFGSETVFMENPGLMILGCHNSDPEN 718
719 QY RGMWALLKVSCKDNTGDDYEDSDIAYLLSKNNAIBPFSQNSRHPSRQKQFNAT 778
719 Db RGMWALLKVSCKDNTGDDYEDSDIAYLLSKNNAIBPFSQNSRHPSRQKQFNAT 778
779 QY TIPENDIEKTDWEAHRTPMKTQNYVSSDLMLLRQS-PTPHGLSLDQAKYETPSD 837
779 Db TIPENDIEKTDWEAHRTPMKTQNYVSSDLMLLRQS-PTPHGLSLDQAKYETPSD 837
838 QY DPSGAIDSNLSMETHPRPOLHSGDMVTFPSGLQLRLNEKLGTTAATLKLDFKV 897
839 Db DPSGAIDSNLSMETHPRPOLHSGDMVTFPSGLQLRLNEKLGTTAATLKLDFKV 897
898 QY SSTSNNLI-STIPSDNLAAGTNTSLGPPSPVHYDSOLDITLPGKXSSPLTESGGPLS 956
899 Db SSTSNNLI-STIPSDNLAAGTNTSLGPPSPVHYDSOLDITLPGKXSSPLTESGGPLS 956
957 QY LSEENNDKLLBSGLMNSQSGWKNVSTESGRIFKGRAGPALLTKDNALFKVSI 1016
959 Db LSEENNDKLLBSGLMNSQSGWKNVSTESGRIFKGRAGPALLTKDNALFKVSI 1016
1017 QY LKNTKTSNNAKXTHIDGSPSLIENSPPWQON-ILESDBTKKVTPIIHDMLMDXNA 1075
1019 Db LKNTKTSNNAKXTHIDGSPSLIENSPPWQON-ILESDBTKKVTPIIHDMLMDXNA 1077
1076 QY TALRLNEMNKTTSSKNMVMQKKEGPTPPDAQNDPMSFFKMLFLPESARMIOPTHGN 1135
1078 Db TALRLNEMNKTTSSKNMVMQKKEGPTPPDAQNDPMSFFKMLFLPESARMIOPTHGN 1137
1136 QY SLNSGGQSPKQVLSLGPCKSVBEGQFLSEKNVYVKGGEFTKDVGLKDMVFPSSRNLEL 1195
1138 Db SLNSGGQSPKQVLSLGPCKSVBEGQFLSEKNVYVKGGEFTKDVGLKDMVFPSSRNLEL 1197
1196 QY TNLNLFENNTNHOEKKIQEBIEKXZTLIQENVVLPOIHTVTGKXNFKNLLSFRQNV 1255
1198 Db TNLNLFENNTNHOEKKIQEBIEKXZTLIQENVVLPOIHTVTGKXNFKNLLSFRQNV 1256
1256 QY EGSYDQAYAPVLQDFRSLNDSINRTKHTAHSK--KGEENLEGLGNQTKQIVERIACT 1313

1257 Db --SLYEVHVPYLQNTSINNSTNVQIHEHFFKRDKETNSEGLVNKTRMVANY--- 1311
1314 QY TRISPNTSOONFVORSKRALQORLPLEETELEKRIIVDDTSTOWSQMKHLIPSTIQT 1373
1312 Db -----PSQKNITQRSKRALQORL-----STQWLKTNCSQCIKQ 1349
1374 QY IDYNEKKGALTOPLSDCLTRSHSIPQANRSLPIAKVSSFPSPRIPLYLVRVLPQDNSS 1433
1350 Db IDESKEMKXFTIKSLSDS-SVIXSTQTNSSDSHIVKTSAPP---PIDLXRSFPQKFS 1405
1434 QY HLPAAZY-----RKXDSGVQESHSFLOQAKONLSAILTLEMTDQREVGSLGTSATNSV 1489
1406 Db HVQASSYIYDFKTSXSRTOESNNFLKETKINNPISLAILEWNNFIDQGFSTPGKSNSTNSV 1465
1490 QY TYKKVENIVLKPDL2KPTSGKVEILLPKVHIYOKDLFPETETNSGSPGHLDLVEGSLQOTE 1549
1466 Db TYKKRENIIFKJPLPEBESGIELLPQVISOEBEILPTEHSGSPGHLNLMKVEFLQIK 1525
1550 QY GAIKWNEANRPGKVPFLRVATESSAKTPSKLLDPLJANDNHYCTQIPIKDEWKSQKSPBKT 1609
1526 Db GPTKWNKAKRHE--SIKGTESKNTESKLLNHHANDYHAAQIPKDMWKSXKSPBII 1583
1610 QY AFKKOTILSLNACESNHAIAINEGONKPEIEVTWAKQGRTERLCSQNPVPLKXHQREI 1669
1584 Db SIKQBDTTLSLRPHGNHSHIGA-NEKQNPQRETIVWQGTQRTCSQIPVPLKXHQREI 1642
1670 QY TRTTLQSQEIBIDYDITISVMKXEDFDIYDEDENQSPRSFOKTRHYFIIAAVERLWDYG 1729
1643 Db --SAFQSQEIBIDYDITISVMKXEDFDIYDEDENQSPRSFOKTRHYFIIAAVERLWDYG 1699
1730 QY MSSSPHVLRNBAQSGSVQFKKXVFQETDGSFTQPLYRGLNEHLGLLGPYIRAEVDN 1789
1700 Db MSSSPHVLRNBAQSGSVQFKKXVFQETDGSFTQPLYRGLNEHLGLLGPYIRAEVDN 1758
1790 QY IMVIFRNOASRPYFSSLSIYBEBORQAGAPRKNFVKPNETKITFYKVOQHMAETKDEF 1849
1759 Db IMVIFRNOASRPYFSSLSIYBEBORQAGAPRKNFVKPNETKITFYKVOQHMAETKDEF 1817
1850 QY DCKAWAYSVDVLEKDVHSLGILPLLVCHTNTLNPAGROVTVQEFALFFITFDTKSWY 1909
1818 Db DCKAWAYSVDVLEKDVHSLGILPLLVCHTNTLNPAGROVTVQEFALFFITFDTKSWY 1877
1910 QY FENNMERNACPNCTOMBDDPTFKRNYRFHANGYIMDTPLGLVMAQDORIRWYLLSMGNN 1969
1878 Db FENNMERNACPNCTOMBDDPTFKRNYRFHANGYIMDTPLGLVMAQDORIRWYLLSMGNN 1937
1970 QY ENIHSHIESGHVFTVRKKEEYKQALYVPGVFTVEMLPKAGIWRVVECLIGEHLHAGM 2029
1938 Db ENIHSHIESGHVFTVRKKEEYKQALYVPGVFTVEMLPKAGIWRVVECLIGEHLHAGM 1997
2030 QY STLFVYVSNKQOTPLGMASGHIRDPQITASGOYGOWAPKLARLHYSGSINAWSTKEPFSW 2089
1998 Db STLFVYVSNKQOTPLGMASGHIRDPQITASGOYGOWAPKLARLHYSGSINAWSTKEPFSW 2057
2090 QY IKVDLLAPMIHGIKTQARQKPSLSIYSQIIMYSLGKQWQTYRGNSTGTLMVFFGNV 2149
2058 Db IKVDLLAPMIHGIKTQARQKPSLSIYSQIIMYSLGKQWQTYRGNSTGTLMVFFGNV 2117
2150 QY DSSGIXHNIENPPIIARVIRLHPTHYSIRSTLRMELMGCDLNSCMPIGMSKALSDAQI 2209
2118 Db DSSGIXHNIENPPIIARVIRLHPTHYSIRSTLRMELMGCDLNSCMPIGMSKALSDAQI 2177
2210 QY TASSYFTNMPATWSPSKARLHQESNAWRPQVNNPKEWLQVDFQTKMKVTGVITQGVKS 2269
2178 Db TASSYFTNMPATWSPSKARLHQESNAWRPQVNNPKEWLQVDFQTKMKVTGVITQGVKS 2237
2270 QY LLSYMYVEFLISSQDGHQWTLFPQNGKVKVQCNQDSFTPVVNSLDPPLLTRVLRHP 2329
2238 Db LLSYMYVEFLISSQDGHQWTLFPQNGKVKVQCNQDSFTPVVNSLDPPLLTRVLRHP 2297
2330 QY QSNVHOIALRMEVLGCEACJLY 2351

```

Db      2298 QIWEHQIALRLIILGCEAQOY 2319
RESULT 14
US-10-187-319-37
; Sequence 37, Application US/10187319
; Publication No. US20030068785A1
; GENERAL INFORMATION:
; APPLICANT: Lollar, John S.
; TITLE OF INVENTION: Hybrid Human/Animal Factor VIII
; NUMBER OF SEQUENCES: 40
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Greenlee, Wilmer and Sullivan, P.C.
; STREET: 5370 Manhattan Circle Suite 201
; CITY: Boulder
; STATE: Colorado
; COUNTRY: USA
; ZIP: 80303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/187,319
; FILING DATE: 27-Aug-2002
; CLASSIFICATION: <unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 09/523,856
; FILING DATE: 2000-03-10
; APPLICATION NUMBER: US 09/037,601
; FILING DATE: 1998-03-10
; APPLICATION NUMBER: WO PCT/US97/11155
; FILING DATE: 1997-06-26
; APPLICATION NUMBER: US 06/670,707
; FILING DATE: 1996-06-26
; ATTORNEY/AGENT INFORMATION:
; NAME: Greenlee, Lorraine L.
; REGISTRATION NUMBER: 27,894
; REFERENCE/DOCKET NUMBER: 75-95K
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 303/499-8080
; TELEFAX: 303/499-8089
; INFORMATION FOR SEQ ID NO: 37:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2133 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 37:
US-10-187-319-37

Query Match      64.9%; Score 8057; DB 14; Length 2133;
Best Local Similarity 56.9%; Pred. No. 0;
Matches 1575; Conserved 209; Mismatches 345; Indels 226; Gaps 17;

QY      1 MQRLSTCFILCLRFCSATRYVILGAVELSDWYQSD-LGELPVDARFPPRPVKSPFF 59
DB      1 MQLELSTCVFLCLLPGLFSATRYVILGAVELSDWYRQSELLRELVDTFFPATAAGALPL 60
QY      60 NTSVYKTKTLFVETFDLPLNIAKAPPPMGLLPTIOAEVYDTVITLKNMASHVPSLHA 119
DB      61 GPSVLKTKTVFVEFTDQLPFSVARPPMVGLLGPTIQAEVYDTVITLKNMASHVPSLHA 120
QY      120 VGVSVYKASEGAEYDDQTSOREKEDDKVPPGSHYVVOVLKENGEMASDPLCLTYSLYS 179
DB      121 VGVSVYKASEGAEYDDQTSOREKEDDKVLPKSKQTVVQVLKENGEMASDPLCLTYSLYS 180
QY      180 HYDLVKDLNSGLIGALLVCRGSLAKEKTQTLHKPILLPAVFDEGKSWHSETKNSLMQDR 239
DB      181 HYDLVKDLNSGLIGALLVCRGSLAKEKTQTLHKPILLPAVFDEGKSWHSETKNSLMQDR 240
QY      240 DAASARAWPMHTVNGYVNRSLPGLIGCRKSKYMEVIGMGTTPVHSLFLEGHTFLVRN 299

```

Db 1211 -GPGPRSAVPRVKOSLKOIRLPLLEBIKPERGVNLNATSTRWS----- 1252
QY 1378 EKEKGAITQSPLSDCLTRSHSIPQANRSPPIAKVSSPFSIRPIVITRVLPQDNSSHLPA 1437
Db 1253 -----BSSPILOGAKNNLSLPFLTLEMAGGKISALGKSAAGPLASGLKLEA 1301
QY 1438 ASYRKDCSGVSSSHFLQAKKNNLSLAILTEMTGDQREVSGLSFATSNSVYTKKVENT 1497
Db 1253 -----BSSPILOGAKNNLSLPFLTLEMAGGKISALGKSAAGPLASGLKLEA 1301
QY 1498 VLPKPDLPKTSKVELLPKHVHYQKDLFTETNSGSPGHLDLVEGSLLOGTGAIKWNEA 1557
Db 1302 VLSAGLSASGKAEPFPKVRHREDLLPQKTSNVSCHGDLGQETFLQKATGPNVNLKV 1361
QY 1558 NRPCKVPLRVATPSSAKTPSKLLDPLAWDNHYGTOTPKHEWKSQKSPKTAFFKKDPT 1617
Db 1362 NRP-----RTPSKULGP-----PMPK-EWESLEKSPKSTALRTKDI 1398
QY 1618 -ISLNAACESNHAIAINEQNKPEIEVTWAKQGRTERLCSQPPVLKRFQREITRTLOS 1676
Db 1399 SLPLDRHESNHSIAAKNEQAEQOREAAWTKQGGPGRLCAPKPPVLRREQRDISLPTROP 1458
QY 1677 DQSEIDYDDTISVEMKEKEDFYDENQSPRSFQKTRHYFIAAVERLWDYGMSSSPHV 1736
Db 1459 EDKMDYDDIFSTETKGEDFYDENQSPRSFQKTRHYFIAAVERLWDYGMSSSPZA 1518
QY 1737 LRNAQSGSVPOPKVVOEFDTGSGTQPLYRGELNEHLGLGPIYRAEVEDNIMVTFN 1796
Db 1519 LRNAQNGEVPRFKVVPREFADGSGTQPSYRGELNKHLLGLGPIYRAEVEDNIMVTFN 1578
QY 1797 QASRPYSYSSLSYSEDOQOQAEPRKPVKPNETXTYFWKVOHMAPTKDEFDCWAY 1856
Db 1579 QASRPYSYSSLSYSPDDQEQAEPRHNFVQPNETRTYFWKQGHMAPTDEPDCWAY 1638
QY 1857 FSDVLEKDVHSGLEPILLVCHTNTNPAHGRVTVQEFALPTIIFDEKSNVFTENMR 1916
Db 1639 FSDVLEKDVHSGLEPILLVCHTNTNPAHGRVTVQEFALPTIIFDEKSNVFTENMR 1698
QY 1917 NCRAPQNIQMEDPTKENTRPHAINGYIMDTLPGLVMAQDQRIWVLLSGNSNENIESTH 1976
Db 1699 NCRAPCHLQVEDPTLXENVPHAINGVYDTPGLVMAQDQRIWVLLSGNSNENIESTH 1758
QY 1977 FSGHVTVRKBEYKVALNLYPGVETVEMLPKSGAGIWRVECLIGEHLHAGNSTLFLVY 2036
Db 1759 FSGHVTVRKBEYKVALNLYPGVETVEMLPKSGAGIWRVECLIGEHLHAGNSTLFLVY 1818
QY 2037 SNKQTEPLGASGHIRDFOITASGOYGOWAPKLARLHYSGSINAWSTKPPSWIKVDLLA 2096
Db 1819 SKBCQPLGASGRIRDFOITASGOYGOWAPKLARLHYSGSINAWSTKPPSWIKVDLLA 1878
QY 2097 PMIIHGKTQGARQKTSLSYISQFIIMYSLDGKKWQTYRGNSTGTLWVFGNVDSGIXH 2156
Db 1879 PMIIHGKTQGARQKTSLSYISQFIIMYSLDGNWQSYRGNSTGTLWVFGNVDSGIXH 1938
QY 2157 NFNPPIIARYIRLHTHTHSIRSLTRELNGCDLNSCMLPGMESKAISDAQITASSYFT 2216
Db 1939 NFNPPIIARYIRLHTHTHSIRSLTRELNGCDLNSCMLPGMESKAISDAQITASSYFT 1998
QY 2217 NMFATWSPKARLHLOGSNWAPQVNNKEWLQVDFQKTVKTVGTTCQVKSLLTSMVY 2276
Db 1999 NMFATWSPKARLHLOGSNWAPQVNNKEWLQVDFQKTVKTVGTTCQVKSLLTSMVY 2058
QY 2277 KEFLISSQDGHQWTLFTQNGKVKVFQGNQDSTFPVNVNSLDPPLLRYLRIHPQSVMHQI 2336
Db 2059 KEFLISSQDGRWTLFLQDGHTKVFGNQDSTFPVNVNSLDPPLLRYLRIHPQSVMHQI 2118
QY 2337 ALRMEVLGCEAQLY 2351
Db 2119 ALRMEVLGCEAQLY 2133

RESULT 15

US-10-131-510A-37

; Sequence 37, Application US/10131510A
; Publication No. US20030166536A1
; GENERAL INFORMATION:
; APPLICANT: Jellar, John S
; TITLE OF INVENTION: Modified Factor VIII
; FILE REFERENCE: 75-95J
; CURRENT APPLICATION NUMBER: US/10/131,510A
; CURRENT FILING DATE: 2002-07-10
; PRIOR FILING DATE: U.S. 09/315,179
; PRIOR FILING DATE: 1999-05-20
; PRIOR APPLICATION NUMBER: U.S. 09/037,601
; PRIOR FILING DATE: 1998-03-10
; PRIOR APPLICATION NUMBER: U.S. 08/670,707
; PRIOR FILING DATE: 1996-06-26
; PRIOR APPLICATION NUMBER: PCT/US97/11155
; PRIOR FILING DATE: 1997-06-26
; PRIOR APPLICATION NUMBER: PCT/US94/13200
; PRIOR FILING DATE: 1994-11-15
; PRIOR APPLICATION NUMBER: U.S. 08/212,133
; PRIOR FILING DATE: 1994-03-11
; PRIOR APPLICATION NUMBER: U.S. 07/864,004
; PRIOR FILING DATE: 1992-04-07
; NUMBER OF SEQ ID NOS: 40
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 37
; LENGTH: 2133
; TYPE: PRT
; ORGANISM: Sus scrofa
US-10-131-510A-37

Query Match 64.9%; Score 8057; DB 14; Length 2133;
Best Local Similarity 66.9%; Pred. No. 0;
Matches 1575; Conservative 209; Mismatches 345; Indels 226; Gaps 17;
QY 1 MQEISTCEFLCLLRFCSATRRYVLGAVELSWDWMQSD-LGELPVDARPPRPVPSPPF 59
Db 1 MQEISTCEFLCLLRFCSATRRYVLGAVELSWDWMQSD-LGELPVDARPPRPVPSPPF 60
QY 60 NTSVVYKKTFLVFETDHLFNIAPKPPPMWGLLPGTIOAEVDTVVTTLKQMAHSPVSLHA 119
Db 61 GFSVLYKKTFLVFETDHLFNIAPKPPPMWGLLPGTIOAEVDTVVTTLKQMAHSPVSLHA 120
QY 120 VGVSVKASGEAYDDQTSQREKEDDVPFGSGHTYVWQVLKENGPMASDPCLTYSYLS 179
Db 121 VGVSVKASGEAYDDQTSQREKEDDVPFGSGHTYVWQVLKENGPMASDPCLTYSYLS 180
QY 180 HYDLVKDLNSGLIGALLVCRSGSLAKETOTLEKILLFAVDEGKSWHSEYNSIMQDR 239
Db 181 HYDLVKDLNSGLIGALLVCRSGSLAKETOTLEKILLFAVDEGKSWHSEYNSIMQDR 240
QY 240 DAASARAMPKMTVANGYVNRSLPGLIGHRKSVYVHVGMTTPEVHSIFLEGHFTLVN 299
Db 241 DPAPAAQAPAMHTVANGYVNRSLPGLIGHRKSVYVHVGMTTPEVHSIFLEGHFTLVN 300
QY 300 HQQASLETSPITFLTAQTLMLDGLQFLSCHISSHQHGMAYKVVDSCEPEFQIMKNN 359
Db 301 HQQASLETSPITFLTAQTLMLDGLQFLSCHISSHQHGMAYKVVDSCEPEFQIMKNN 360
QY 360 EBAEDYDDDLTSEMDVVRFDONSPSFQIHSVAKKPKTWVHYIAAEEEDWDYAPLV 419
Db 361 EBAEDYDDDLTSEMDVVRFDONSPSFQIHSVAKKPKTWVHYIAAEEEDWDYAPLV 419
QY 420 APDRSYKSOYLNNGPQIRGRYKVRFMAYVDETFTKREAIQHESGILGPLLYGEVGT 479
Db 420 APDRSYKSOYLNNGPQIRGRYKVRFMAYVDETFTKREAIQHESGILGPLLYGEVGT 479
QY 480 LLIIIPKQASRPYNYPHGITTIVRPLYSRRLPKGVKHLKDFPILPCEIFKYYKWTVDEG 539
Db 480 LLIIIPKQASRPYNYPHGITTIVRPLYSRRLPKGVKHLKDFPILPCEIFKYYKWTVDEG 539
QY 540 PTKSDPRCLTRYKSSPVMNERDLASGLIGPLICYKESVDQKGNQIMSKNVILFSVD 599
Db 540 PTKSDPRCLTRYKSSPVMNERDLASGLIGPLICYKESVDQKGNQIMSKNVILFSVD 599

[illegible]

Search completed: April 13, 2004, 14:20:04
Job time : 74 secs

This Page Blank (uspto)

This Page Blank (uspto)

This Page Blank (uspto)